

US011302116B2

(12) **United States Patent**  
**Glaser et al.**

(10) **Patent No.:** **US 11,302,116 B2**  
(45) **Date of Patent:** **\*Apr. 12, 2022**

(54) **DEVICE INTERFACE EXTRACTION**

5/2257 (2013.01); *H04N 7/18* (2013.01);  
*H04N 7/183* (2013.01); *H05B 47/105*  
(2020.01); *H05B 47/125* (2020.01); *H05B*  
*47/19* (2020.01); *G06V 20/63* (2022.01)

(71) Applicant: **Grabango Co.**, Berkeley, CA (US)

(72) Inventors: **William Glaser**, Berkeley, CA (US);  
**Brian Van Osdol**, Piedmont, CA (US)

(58) **Field of Classification Search**

None  
See application file for complete search history.

(73) Assignee: **Grabango Co.**, Berkeley, CA (US)

(\*) Notice: Subject to any disclaimer, the term of this  
patent is extended or adjusted under 35  
U.S.C. 154(b) by 40 days.

This patent is subject to a terminal dis-  
claimer.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,254,853 A 10/1993 Reich  
5,418,567 A 5/1995 Boers et al.  
(Continued)

(21) Appl. No.: **16/803,419**

FOREIGN PATENT DOCUMENTS

(22) Filed: **Feb. 27, 2020**

BR 102016019151 A2 3/2018  
WO 2019032304 A1 2/2019

(65) **Prior Publication Data**

US 2020/0204391 A1 Jun. 25, 2020

OTHER PUBLICATIONS

Elizabeth Weise: "How Amazon's line-less grocery services might  
really work", USA Today, Dec. 8, 2016 (Dec. 8, 2016), Retrieved  
from the Internet: URL: <https://eu.usatoday.com/story/tech/news/2016/12/06/amazon-go-surveillance-cameras-shopping-grocery-supermarket/95055200/> [retrieved on Feb. 21, 2020] \*the whole  
document\*.

**Related U.S. Application Data**

(63) Continuation of application No. 15/644,799, filed on  
Jul. 9, 2017, now Pat. No. 10,615,994.  
(Continued)

(51) **Int. Cl.**

*G06V 40/20* (2022.01)  
*G08B 13/196* (2006.01)

(Continued)

*Primary Examiner* — Delomia L Gilliard

(74) *Attorney, Agent, or Firm* — Alpine Patents LLC;  
Brian Van Osdol

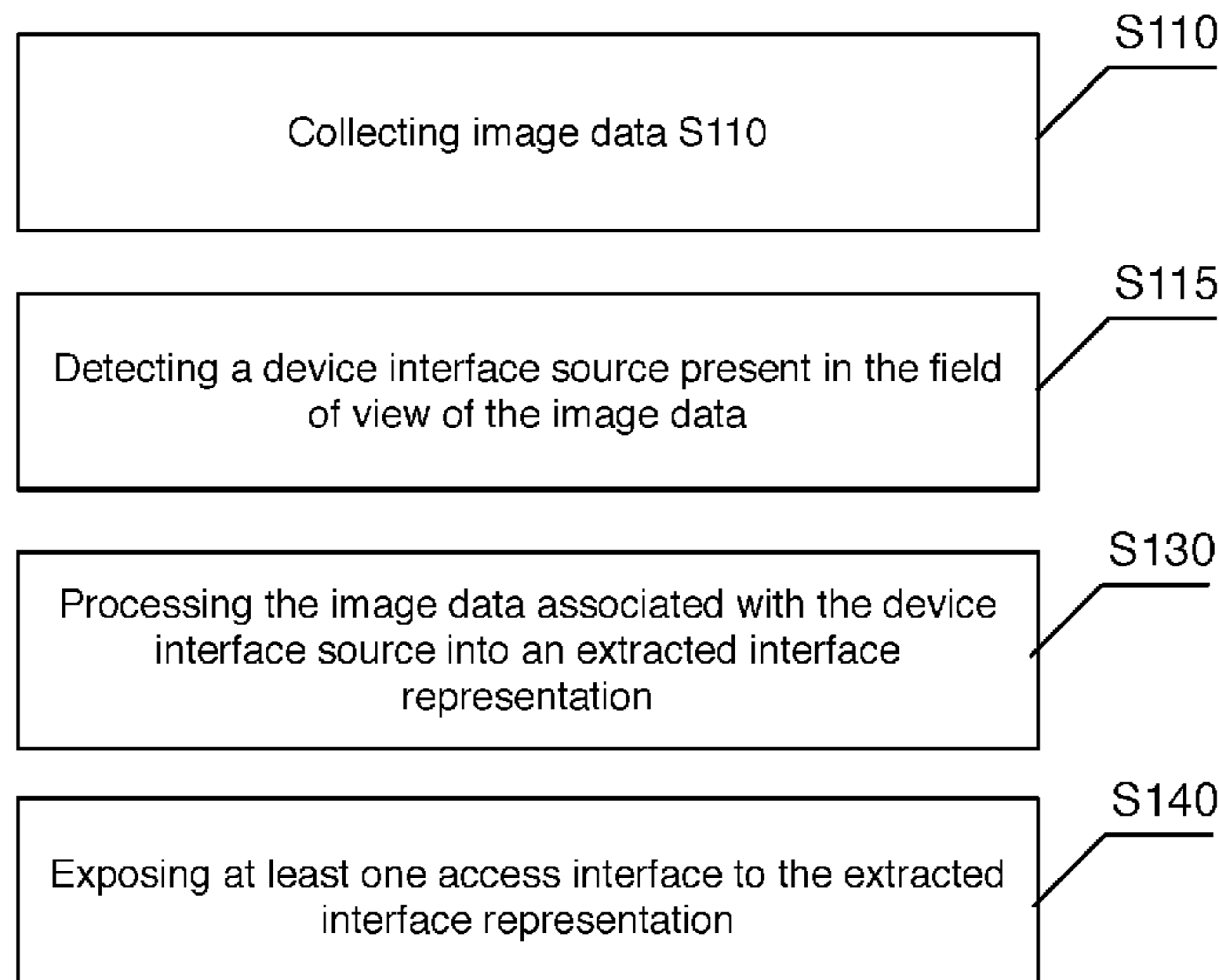
(52) **U.S. Cl.**

CPC ..... *G06V 40/20* (2022.01); *G06K 9/6232*  
(2013.01); *G06V 10/245* (2022.01); *G06V*  
*10/40* (2022.01); *G06V 20/00* (2022.01);  
*G06V 20/52* (2022.01); *G06V 40/10*  
(2022.01); *G08B 13/196* (2013.01); *H04L*  
*12/282* (2013.01); *H04L 12/2809* (2013.01);  
*H04L 12/2814* (2013.01); *H04L 67/12*  
(2013.01); *H04N 5/2252* (2013.01); *H04N*

(57) **ABSTRACT**

A system and method for visually automated interface  
integration that includes collecting image data; detecting a  
device interface source in the image data; processing the  
image data associated with the device interface source into  
an extracted interface representation; and exposing at least  
one access interface to the extracted interface representation.

**19 Claims, 39 Drawing Sheets**



**Related U.S. Application Data**

(60) Provisional application No. 62/360,366, filed on Jul. 9, 2016, provisional application No. 62/360,369, filed on Jul. 9, 2016.

(51) **Int. Cl.**

*H04N 5/225* (2006.01)  
*H04N 7/18* (2006.01)  
*H05B 47/19* (2020.01)  
*H05B 47/125* (2020.01)  
*H05B 47/105* (2020.01)  
*G06V 10/40* (2022.01)  
*G06V 10/24* (2022.01)  
*G06V 20/00* (2022.01)  
*G06V 20/52* (2022.01)  
*G06V 40/10* (2022.01)  
*G06K 9/62* (2022.01)  
*H04L 67/12* (2022.01)  
*H04L 12/28* (2006.01)  
*G06V 20/62* (2022.01)

9,474,934 B1 10/2016 Krueger et al.  
 9,740,977 B1 8/2017 Moon et al.  
 9,747,497 B1 8/2017 Sharma et al.  
 9,892,438 B1 2/2018 Kundu et al.  
 9,911,290 B1 3/2018 Zalewski et al.  
 9,948,902 B1 4/2018 Trundle  
 10,055,853 B1 8/2018 Fisher et al.  
 10,083,358 B1 9/2018 Shin et al.  
 10,127,438 B1 11/2018 Fisher et al.  
 10,133,933 B1 11/2018 Fisher et al.  
 10,134,004 B1 11/2018 Liberato, Jr. et al.  
 10,198,080 B1 2/2019 Worley, III et al.  
 10,198,625 B1 2/2019 Shin et al.  
 10,217,120 B1 2/2019 Shin et al.  
 10,262,331 B1 4/2019 Sharma et al.  
 10,282,621 B2 5/2019 Glaser et al.  
 10,296,936 B1 5/2019 Saurabh et al.  
 10,339,595 B2 7/2019 Glaser et al.  
 10,347,009 B1 7/2019 Terven et al.  
 10,354,262 B1 7/2019 Hershey et al.  
 10,387,896 B1 8/2019 Hershey et al.  
 10,474,858 B2 11/2019 Davis et al.  
 10,659,247 B2 5/2020 Glaser et al.  
 2001/0010541 A1 8/2001 Fernandez et al.  
 2002/0122559 A1 9/2002 Fay et al.  
 2003/0210340 A1 11/2003 Romanowich  
 2003/0216969 A1 11/2003 Bauer et al.  
 2004/0201754 A1 10/2004 McAlister  
 2004/0260513 A1 12/2004 Fitzpatrick et al.  
 2005/0096997 A1 5/2005 Jain et al.  
 2005/0189411 A1 9/2005 Ostrowski et al.  
 2006/0032915 A1 2/2006 Schwartz  
 2006/0147087 A1 7/2006 Goncalves et al.  
 2007/0091177 A1 4/2007 West et al.  
 2007/0146873 A1 6/2007 Ortyl et al.  
 2007/0242300 A1 10/2007 Inai  
 2007/0284440 A1 12/2007 Birmingham et al.  
 2008/0226129 A1 9/2008 Kundu et al.  
 2008/0228585 A1 9/2008 Duri et al.  
 2009/0195648 A1 8/2009 Thomas et al.  
 2010/0020173 A1 1/2010 Marquart et al.  
 2010/0255902 A1 10/2010 Goldstein et al.  
 2010/0280956 A1 11/2010 Chutorash et al.  
 2010/0295946 A1 11/2010 Reed et al.  
 2011/0122231 A1 5/2011 Fujieda et al.  
 2011/0215147 A1 9/2011 Goncalves  
 2012/0019168 A1 1/2012 Noda et al.  
 2012/0027297 A1 2/2012 Feris et al.  
 2012/0173351 A1 7/2012 Hanson et al.  
 2013/0103537 A1 4/2013 Hewett  
 2013/0147963 A1 6/2013 Henninger et al.  
 2013/0177201 A1 7/2013 Fisher  
 2013/0183042 A1 7/2013 Knapp et al.  
 2013/0215116 A1 8/2013 Siddique et al.  
 2013/0223673 A1 8/2013 Davis et al.  
 2013/0284806 A1 10/2013 Margalit  
 2013/0290557 A1 10/2013 Baratz  
 2013/0317300 A1 11/2013 Berci et al.  
 2013/0335571 A1 12/2013 Libal  
 2013/0342688 A1 12/2013 Siu  
 2014/0082519 A1 3/2014 Wang et al.  
 2014/0082610 A1 3/2014 Wang et al.  
 2014/0129688 A1 5/2014 Asenjo et al.  
 2014/0245160 A1 8/2014 Bauer et al.  
 2014/0263631 A1 9/2014 Muniz  
 2014/0265880 A1 9/2014 Taipale et al.  
 2014/0272855 A1 9/2014 Maser et al.  
 2014/0279191 A1 9/2014 Agarwal et al.  
 2014/0285660 A1 9/2014 Jamtgaard et al.  
 2014/0297487 A1 10/2014 Bashkin  
 2014/0330408 A1 11/2014 Rolley  
 2014/0363059 A1 12/2014 Hurewitz  
 2014/0365334 A1 12/2014 Hurewitz  
 2014/0372957 A1 12/2014 Keane et al.  
 2015/0012396 A1 1/2015 Puerini et al.  
 2015/0019391 A1 1/2015 Kumar et al.  
 2015/0029339 A1 1/2015 Kobres et al.  
 2015/0039458 A1 2/2015 Reid  
 2015/0046213 A1 2/2015 Doreswamy et al.

(56)

**References Cited**

U.S. PATENT DOCUMENTS

5,502,564 A 3/1996 Ledger  
 6,970,183 B1 11/2005 Monroe  
 6,996,460 B1 2/2006 Krahnstoever et al.  
 7,053,915 B1 5/2006 Jung et al.  
 7,195,157 B2 3/2007 Swartz et al.  
 7,225,414 B1 5/2007 Sharma et al.  
 7,227,976 B1 6/2007 Jung et al.  
 7,274,803 B1 9/2007 Sharma et al.  
 7,283,650 B1 10/2007 Sharma et al.  
 7,319,779 B1 1/2008 Mummareddy et al.  
 7,505,621 B1 3/2009 Agrawal et al.  
 7,590,261 B1 9/2009 Mariano et al.  
 7,711,155 B1 5/2010 Sharma et al.  
 7,734,070 B1 6/2010 Sharma et al.  
 7,848,548 B1 12/2010 Moon et al.  
 7,911,482 B1 3/2011 Mariano et al.  
 7,912,246 B1 3/2011 Moon et al.  
 7,921,036 B1 4/2011 Sharma et al.  
 7,930,204 B1 4/2011 Sharma et al.  
 7,957,565 B1 6/2011 Sharma et al.  
 7,974,869 B1 7/2011 Sharma et al.  
 7,987,111 B1 7/2011 Sharma et al.  
 8,009,863 B1 8/2011 Sharma et al.  
 8,010,402 B1 8/2011 Sharma et al.  
 8,027,521 B1 9/2011 Moon et al.  
 8,098,888 B1 1/2012 Mummareddy et al.  
 8,189,926 B2 5/2012 Sharma et al.  
 8,219,438 B1 7/2012 Moon et al.  
 8,254,633 B1 8/2012 Moon et al.  
 8,295,597 B1 10/2012 Sharma et al.  
 8,325,982 B1 12/2012 Moon et al.  
 8,351,647 B2 1/2013 Sharma et al.  
 8,379,937 B1 2/2013 Moon et al.  
 8,380,558 B1 2/2013 Sharma et al.  
 8,396,758 B2 3/2013 Paradise et al.  
 8,412,656 B1 4/2013 Baboo et al.  
 8,433,612 B1 4/2013 Sharma et al.  
 8,448,859 B2 5/2013 Goncalves et al.  
 8,520,906 B1 8/2013 Moon et al.  
 8,577,705 B1 11/2013 Baboo et al.  
 8,812,344 B1 8/2014 Saurabh et al.  
 9,120,621 B1 9/2015 Curlander et al.  
 9,141,931 B2 9/2015 Ackerman  
 9,161,084 B1 10/2015 Sharma et al.  
 9,262,681 B1 2/2016 Mishra  
 9,270,634 B1 2/2016 Gu et al.  
 9,317,785 B1 4/2016 Moon et al.  
 9,412,099 B1 8/2016 Tyree  
 9,426,720 B2 8/2016 Cohn et al.

2004/0201754 A1 10/2004 McAlister  
 2004/0260513 A1 12/2004 Fitzpatrick et al.  
 2005/0096997 A1 5/2005 Jain et al.  
 2005/0189411 A1 9/2005 Ostrowski et al.  
 2006/0032915 A1 2/2006 Schwartz  
 2006/0147087 A1 7/2006 Goncalves et al.  
 2007/0091177 A1 4/2007 West et al.  
 2007/0146873 A1 6/2007 Ortyl et al.  
 2007/0242300 A1 10/2007 Inai  
 2007/0284440 A1 12/2007 Birmingham et al.  
 2008/0226129 A1 9/2008 Kundu et al.  
 2008/0228585 A1 9/2008 Duri et al.  
 2009/0195648 A1 8/2009 Thomas et al.  
 2010/0020173 A1 1/2010 Marquart et al.  
 2010/0255902 A1 10/2010 Goldstein et al.  
 2010/0280956 A1 11/2010 Chutorash et al.  
 2010/0295946 A1 11/2010 Reed et al.  
 2011/0122231 A1 5/2011 Fujieda et al.  
 2011/0215147 A1 9/2011 Goncalves  
 2012/0019168 A1 1/2012 Noda et al.  
 2012/0027297 A1 2/2012 Feris et al.  
 2012/0173351 A1 7/2012 Hanson et al.  
 2013/0103537 A1 4/2013 Hewett  
 2013/0147963 A1 6/2013 Henninger et al.  
 2013/0177201 A1 7/2013 Fisher  
 2013/0183042 A1 7/2013 Knapp et al.  
 2013/0215116 A1 8/2013 Siddique et al.  
 2013/0223673 A1 8/2013 Davis et al.  
 2013/0284806 A1 10/2013 Margalit  
 2013/0290557 A1 10/2013 Baratz  
 2013/0317300 A1 11/2013 Berci et al.  
 2013/0335571 A1 12/2013 Libal  
 2013/0342688 A1 12/2013 Siu  
 2014/0082519 A1 3/2014 Wang et al.  
 2014/0082610 A1 3/2014 Wang et al.  
 2014/0129688 A1 5/2014 Asenjo et al.  
 2014/0245160 A1 8/2014 Bauer et al.  
 2014/0263631 A1 9/2014 Muniz  
 2014/0265880 A1 9/2014 Taipale et al.  
 2014/0272855 A1 9/2014 Maser et al.  
 2014/0279191 A1 9/2014 Agarwal et al.  
 2014/0285660 A1 9/2014 Jamtgaard et al.  
 2014/0297487 A1 10/2014 Bashkin  
 2014/0330408 A1 11/2014 Rolley  
 2014/0363059 A1 12/2014 Hurewitz  
 2014/0365334 A1 12/2014 Hurewitz  
 2014/0372957 A1 12/2014 Keane et al.  
 2015/0012396 A1 1/2015 Puerini et al.  
 2015/0019391 A1 1/2015 Kumar et al.  
 2015/0029339 A1 1/2015 Kobres et al.  
 2015/0039458 A1 2/2015 Reid  
 2015/0046213 A1 2/2015 Doreswamy et al.



(56)

References Cited

U.S. PATENT DOCUMENTS

2015/0077787 A1	3/2015	Nishimura et al.	2017/0131781 A1	5/2017	Buban
2015/0077797 A1	3/2015	Kurokawa	2017/0161703 A1	6/2017	Dodia
2015/0088694 A1	3/2015	Ackerman	2017/0169440 A1	6/2017	Dey et al.
2015/0095189 A1	4/2015	Dharssi et al.	2017/0178352 A1	6/2017	Harmsen et al.
2015/0097961 A1	4/2015	Ure et al.	2017/0188013 A1	6/2017	Presler
2015/0124973 A1	5/2015	Arteaga et al.	2017/0216667 A1	8/2017	Garvey et al.
2015/0133190 A1	5/2015	Fisher et al.	2017/0278175 A1	9/2017	Park et al.
2015/0138383 A1	5/2015	Kelley et al.	2017/0316656 A1	11/2017	Chaubard et al.
2015/0154973 A1	6/2015	McKenna et al.	2017/0332956 A1	11/2017	Bigolin et al.
2015/0156332 A1	6/2015	Kandregula	2018/0005044 A1*	1/2018	Olson ..... H04N 5/2257
2015/0163412 A1	6/2015	Holley et al.	2018/0012072 A1	1/2018	Glaser et al.
2015/0227890 A1	8/2015	Bednarek et al.	2018/0012080 A1*	1/2018	Glaser ..... G06K 9/46
2015/0244992 A1	8/2015	Buehler	2018/0018508 A1	1/2018	Tusch
2015/0282282 A1	10/2015	Breuer et al.	2018/0107968 A1	4/2018	Wu et al.
2015/0310601 A1	10/2015	Rodriguez et al.	2018/0189763 A1	7/2018	Olmstead et al.
2015/0327045 A1	11/2015	Chang et al.	2018/0232796 A1	8/2018	Glaser et al.
2015/0373509 A1	12/2015	Wang et al.	2018/0240180 A1	8/2018	Glaser et al.
2016/0012379 A1	1/2016	Iwai	2018/0245736 A1	8/2018	Patel
2016/0019514 A1	1/2016	Landers et al.	2018/0300553 A1	10/2018	Khosla et al.
2016/0027262 A1	1/2016	Skotty et al.	2018/0322209 A1	11/2018	Jin et al.
2016/0037135 A1	2/2016	McSheffrey et al.	2018/0332235 A1	11/2018	Glaser
2016/0100086 A1	4/2016	Chien	2018/0332236 A1	11/2018	Glaser et al.
2016/0110791 A1	4/2016	Herring et al.	2018/0373928 A1	12/2018	Glaser et al.
2016/0112608 A1	4/2016	Elensi et al.	2019/0005479 A1	1/2019	Glaser et al.
2016/0132854 A1	5/2016	Singh	2019/0028643 A1	1/2019	Oryoji
2016/0173827 A1	6/2016	Dannan et al.	2019/0043003 A1	2/2019	Fisher et al.
2016/0217388 A1	7/2016	Okanohara et al.	2019/0054347 A1	2/2019	Saigh et al.
2016/0224856 A1	8/2016	Park et al.	2019/0079591 A1	3/2019	Glaser et al.
2016/0242252 A1	8/2016	Lim et al.	2019/0114488 A1	4/2019	Glaser et al.
2016/0254864 A1	9/2016	Mueller et al.	2019/0116322 A1	4/2019	Holzer et al.
2016/0270191 A1	9/2016	Ludwig, Jr. et al.	2019/0156273 A1	5/2019	Fisher et al.
2016/0282039 A1	9/2016	Motukuri et al.	2019/0156274 A1	5/2019	Fisher et al.
2016/0289964 A1	10/2016	Engberg	2019/0156275 A1	5/2019	Fisher et al.
2016/0321506 A1	11/2016	Fridental et al.	2019/0156276 A1	5/2019	Fisher et al.
2016/0345414 A1	11/2016	Nolan et al.	2019/0156277 A1	5/2019	Fisher et al.
2016/0358312 A1	12/2016	Kolb et al.	2019/0156506 A1	5/2019	Fisher et al.
2016/0379074 A1	12/2016	Nielsen et al.	2019/0205933 A1	7/2019	Glaser et al.
2017/0032182 A1	2/2017	Motukuri et al.	2019/0253270 A1	8/2019	Glaser et al.
2017/0039613 A1	2/2017	Kaehler et al.	2019/0333039 A1	10/2019	Glaser et al.
2017/0053171 A1	2/2017	Buehler	2020/0079412 A1	3/2020	Ramanathan et al.
2017/0108236 A1	4/2017	Guan et al.	2020/0134590 A1	4/2020	Glaser et al.
2017/0123030 A1	5/2017	Hengerer et al.	2020/0160670 A1	5/2020	Zalewski et al.
			2020/0265494 A1	8/2020	Glaser et al.

\* cited by examiner

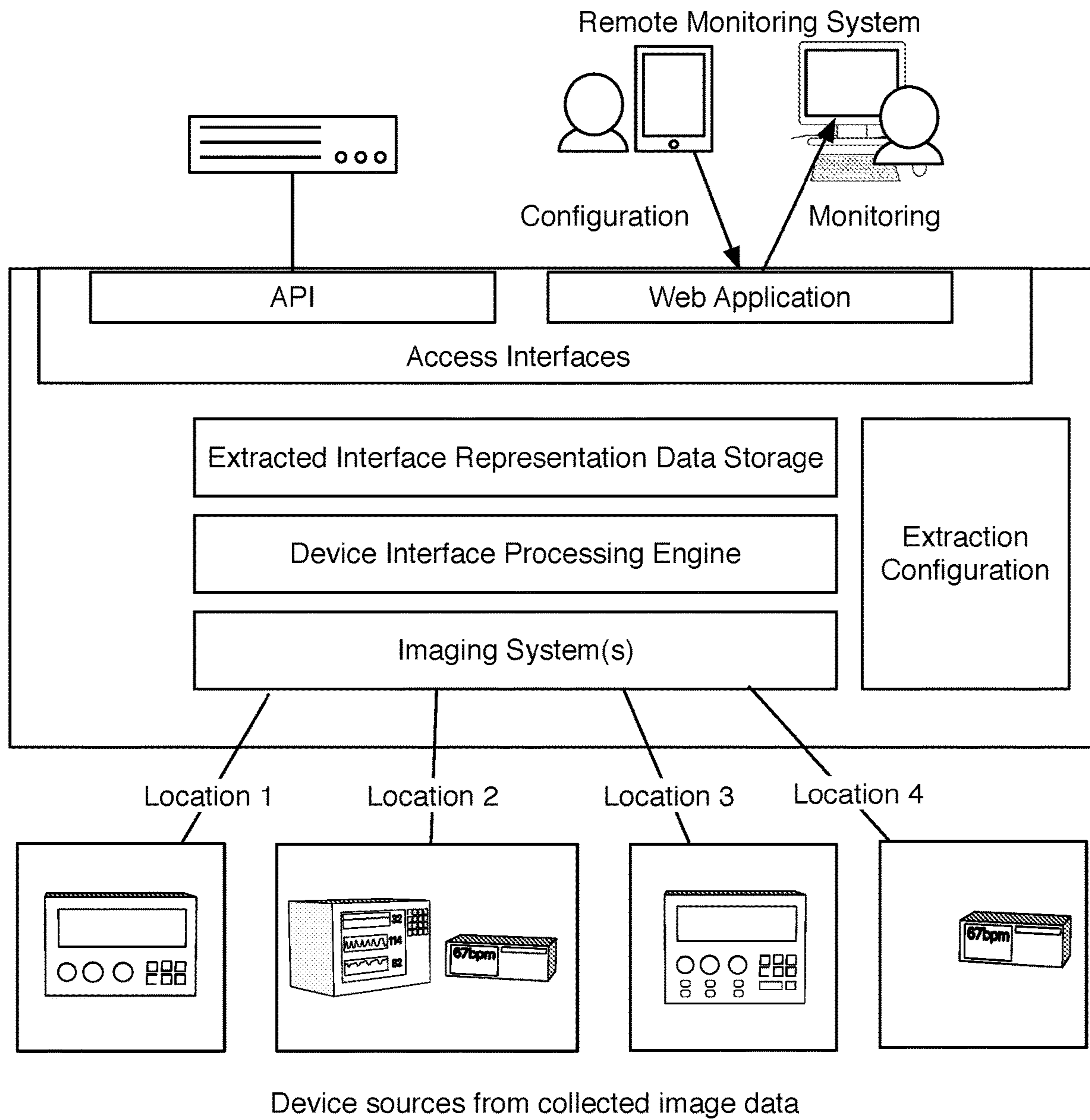


FIGURE 1

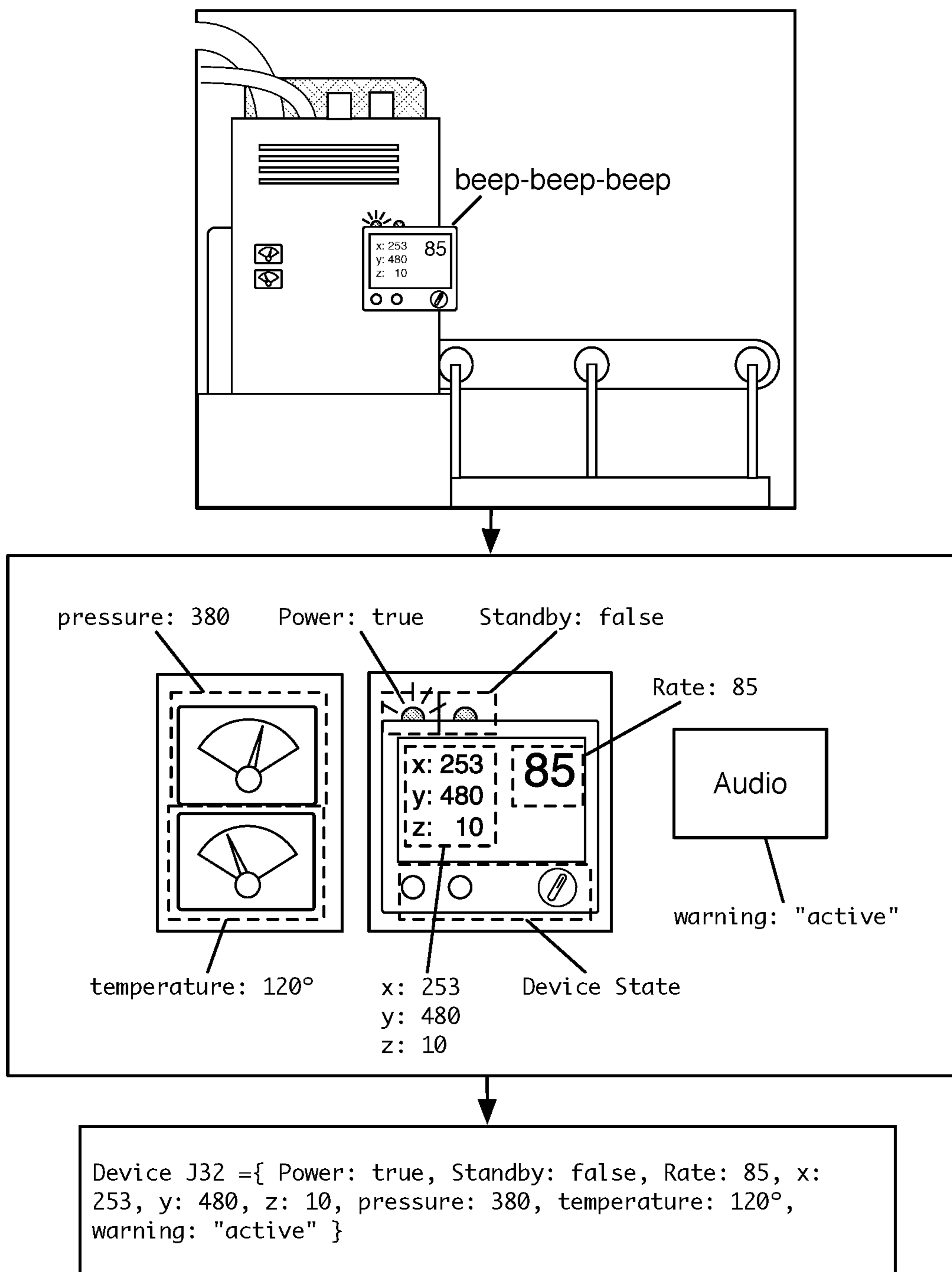


FIGURE 2

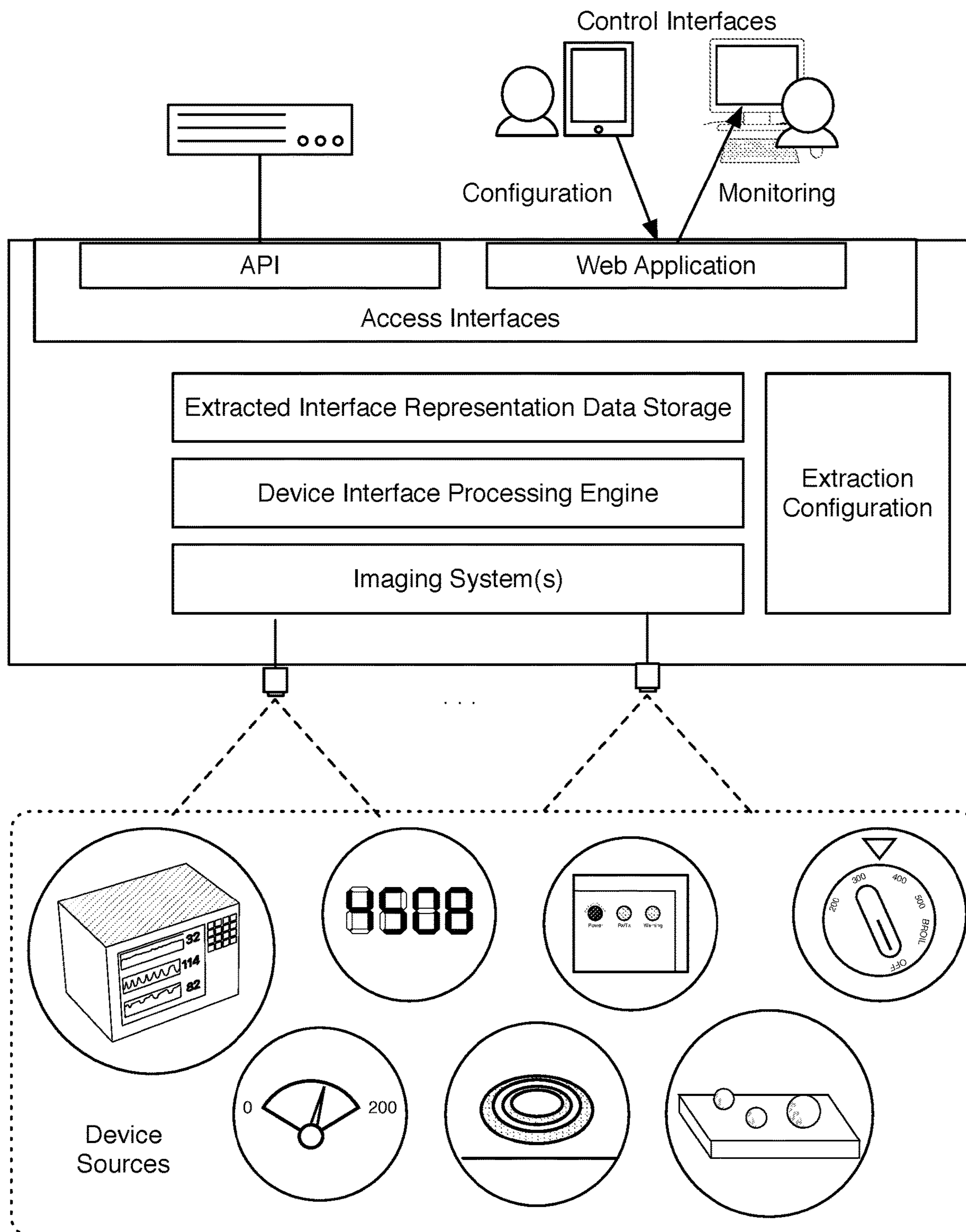
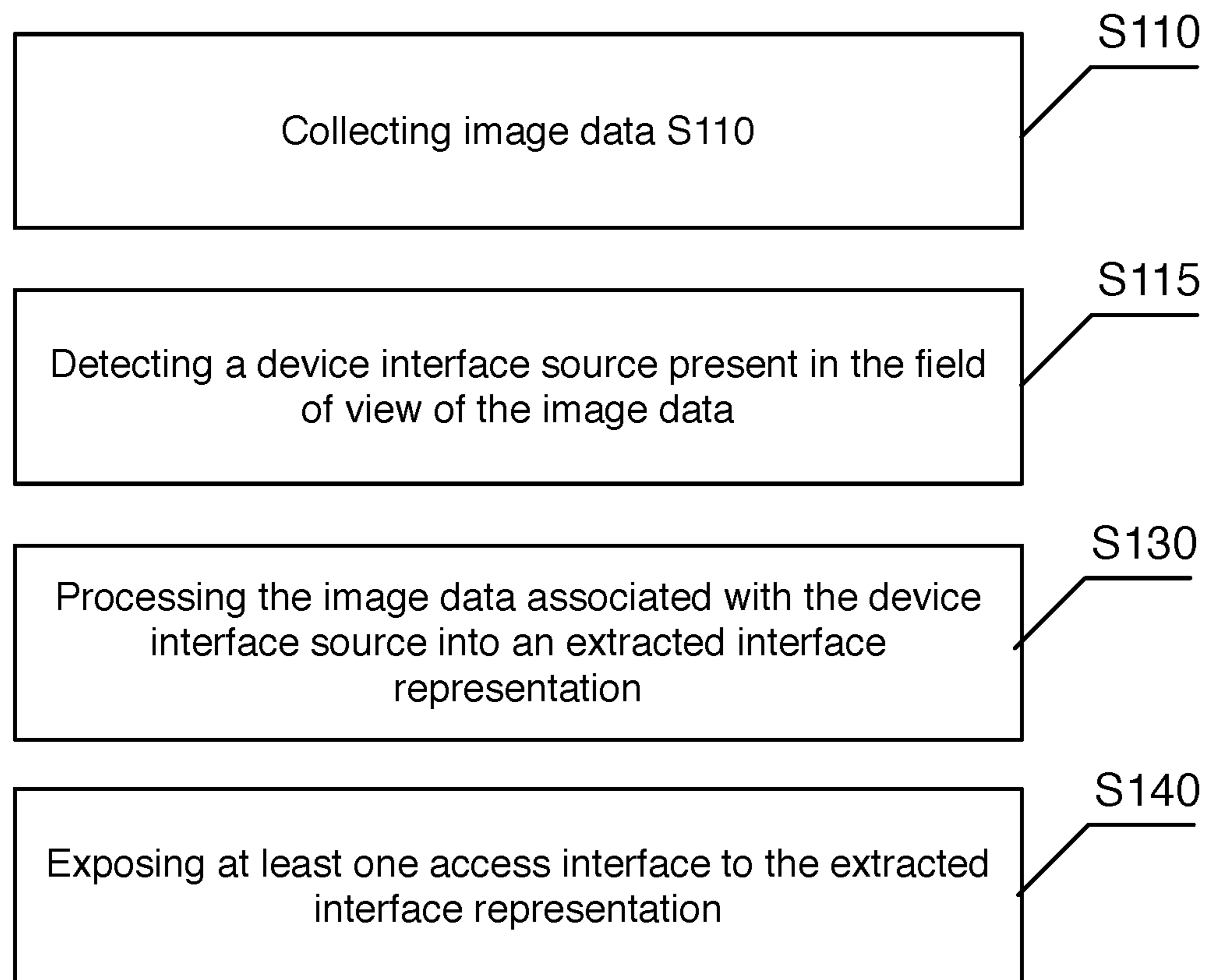
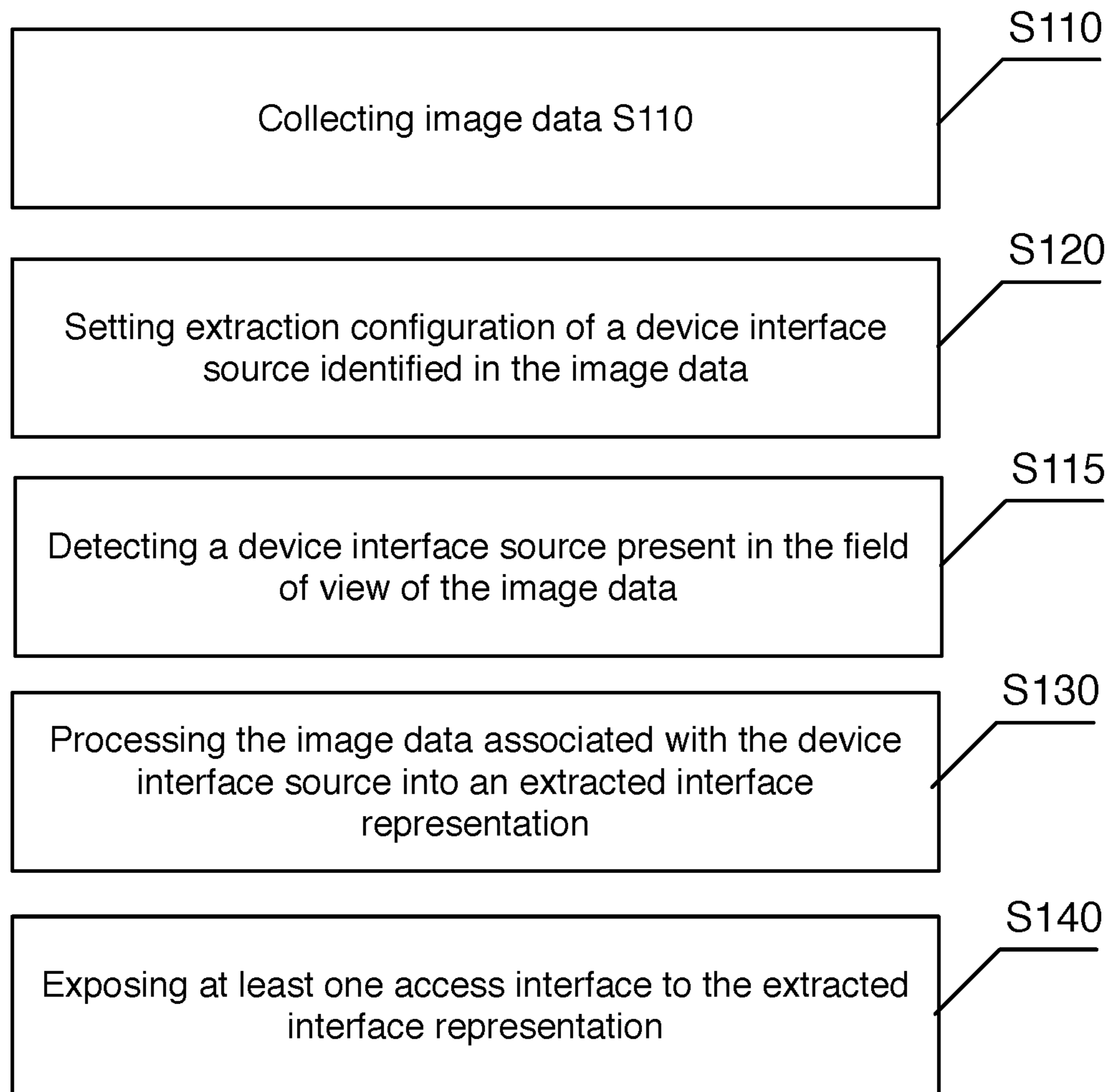


FIGURE 3



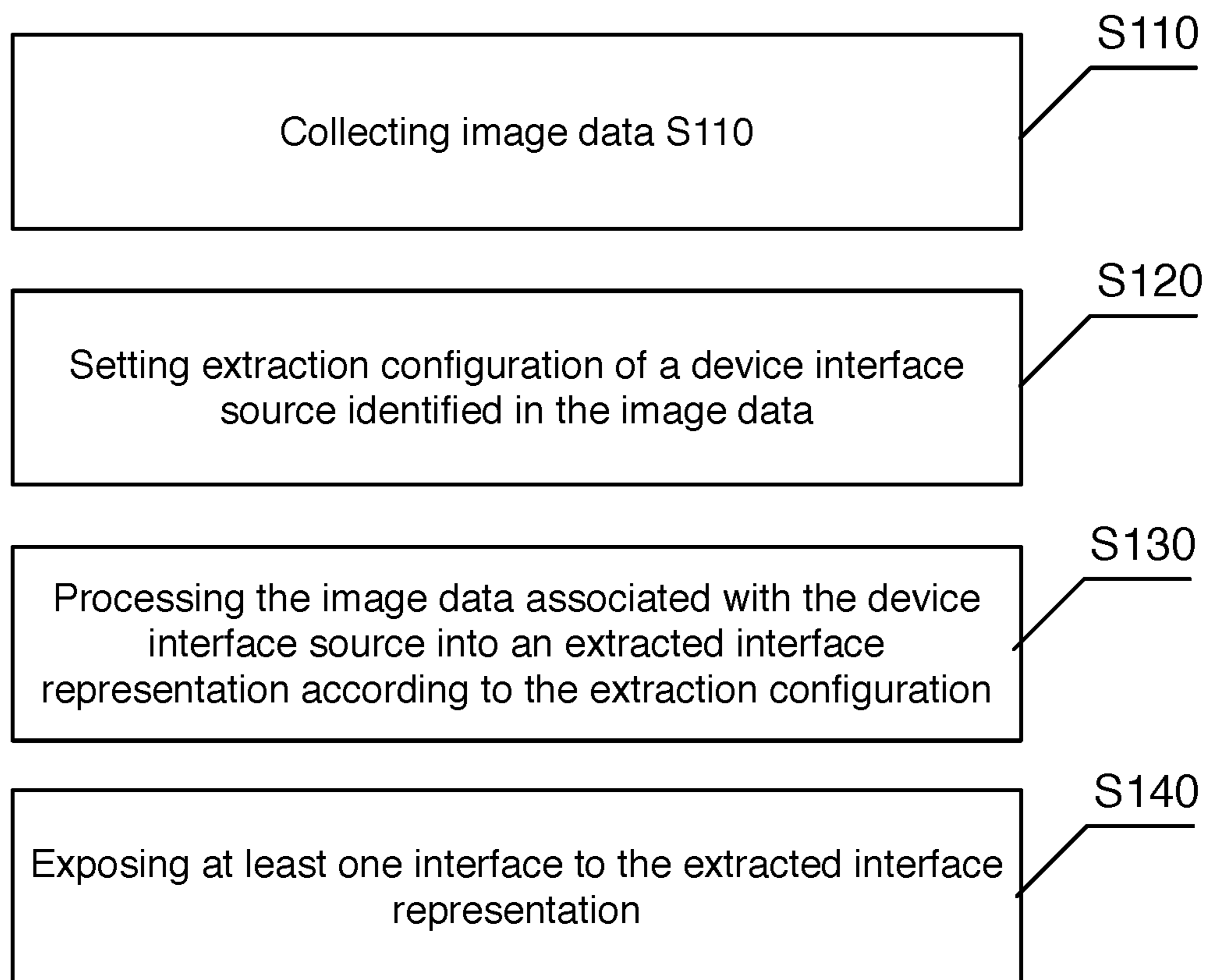


**FIGURE 4**



**FIGURE 5**





**FIGURE 6**

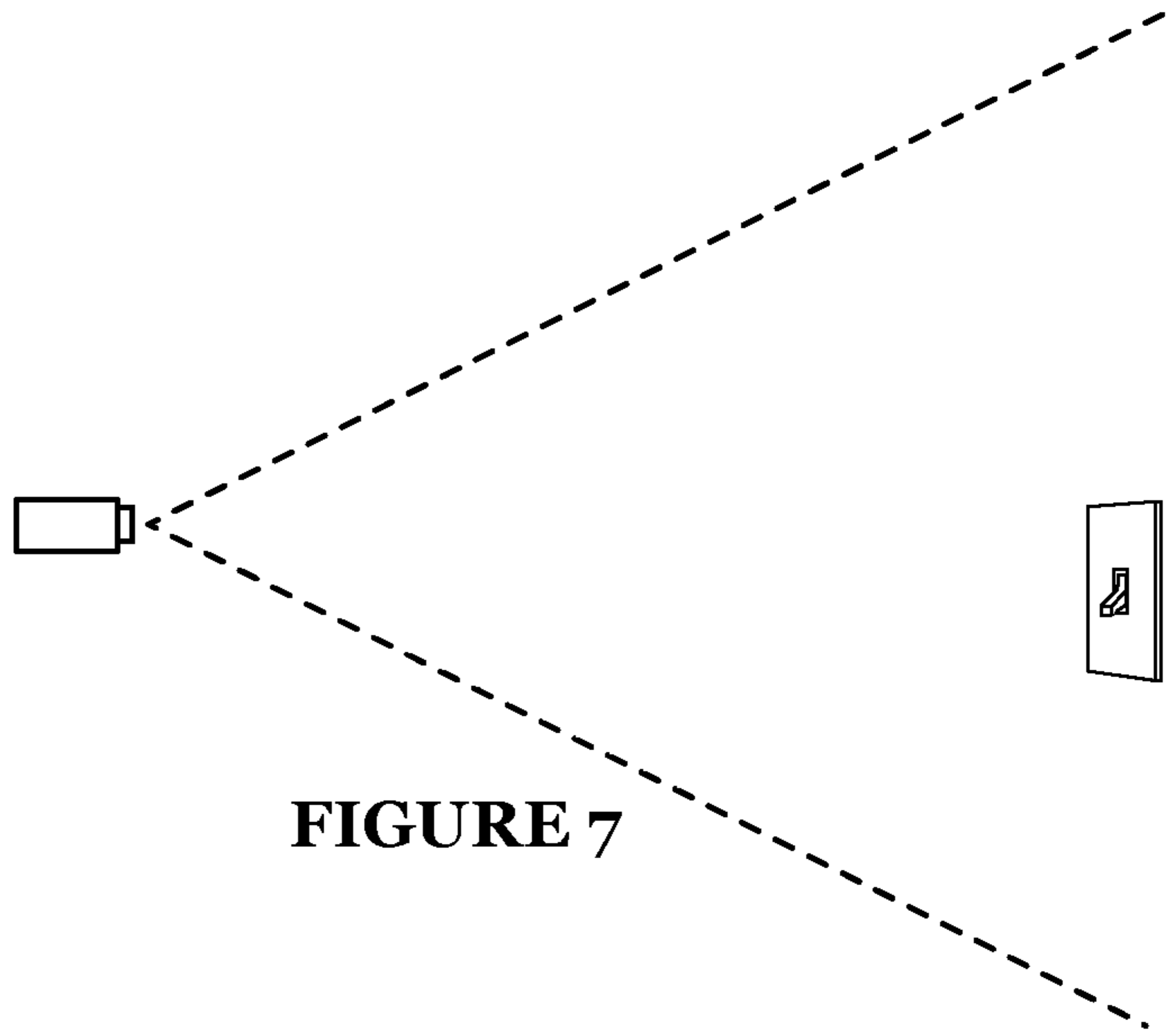


FIGURE 7

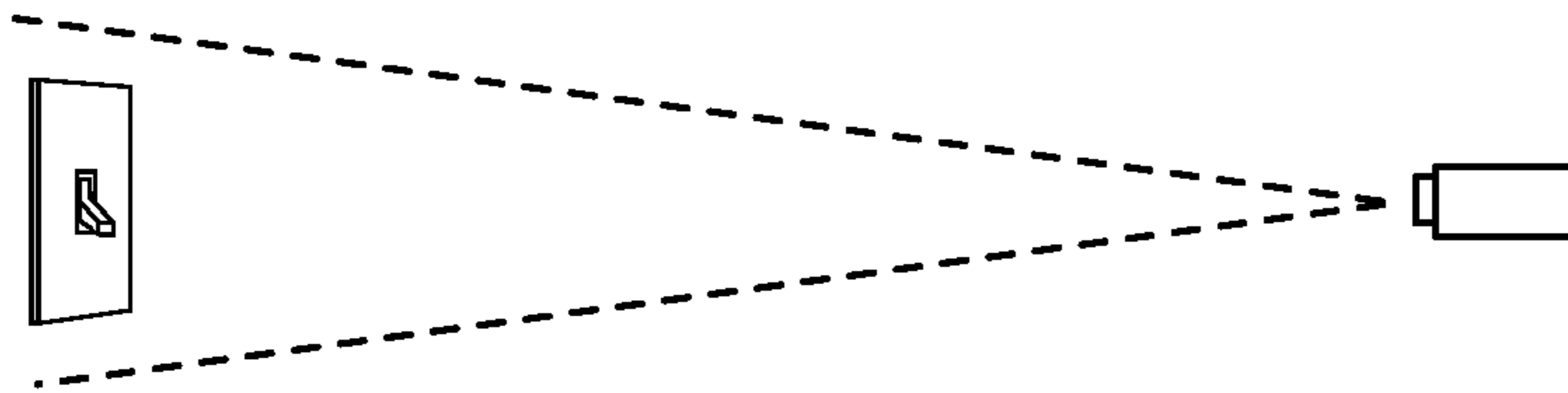


FIGURE 8

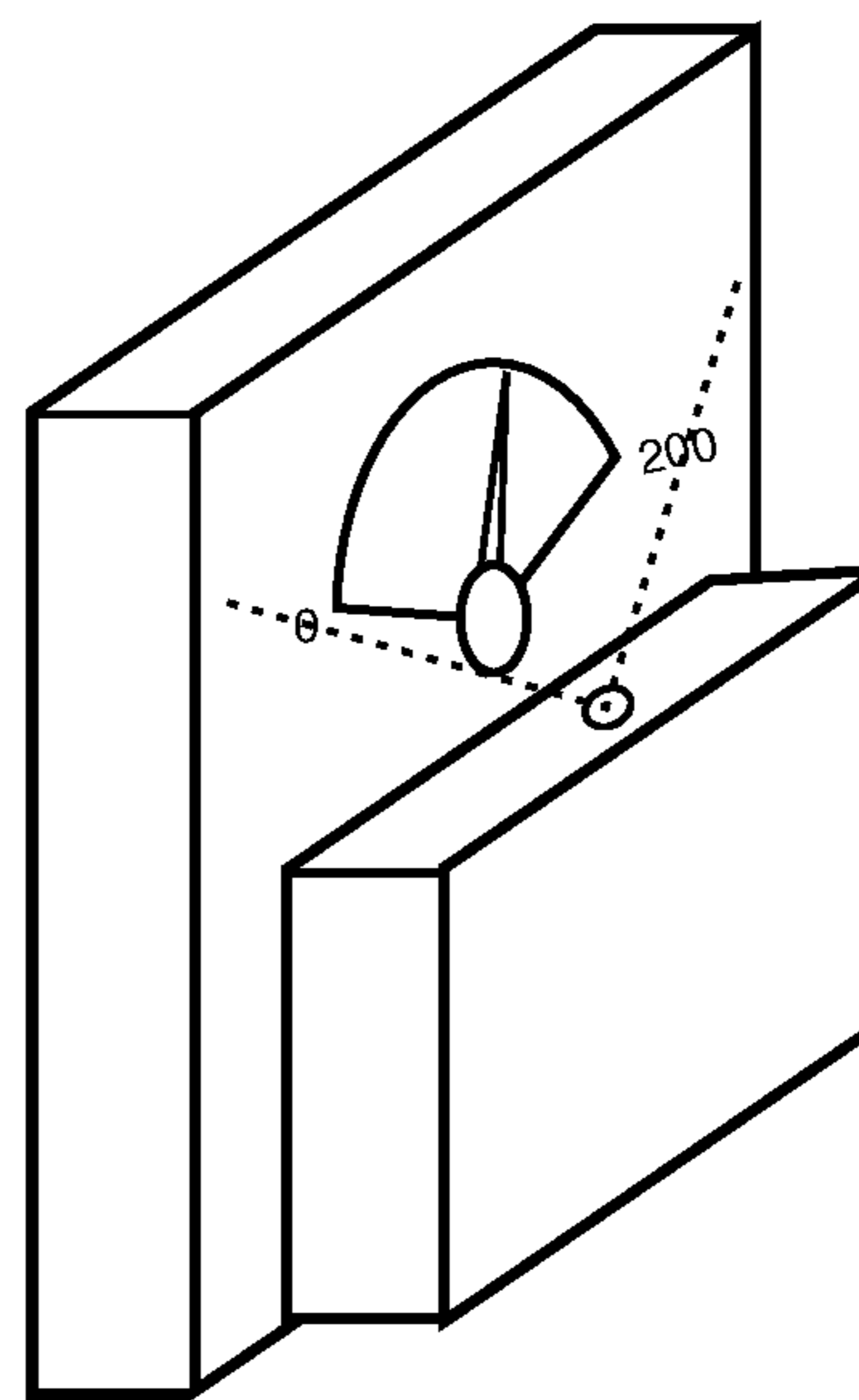
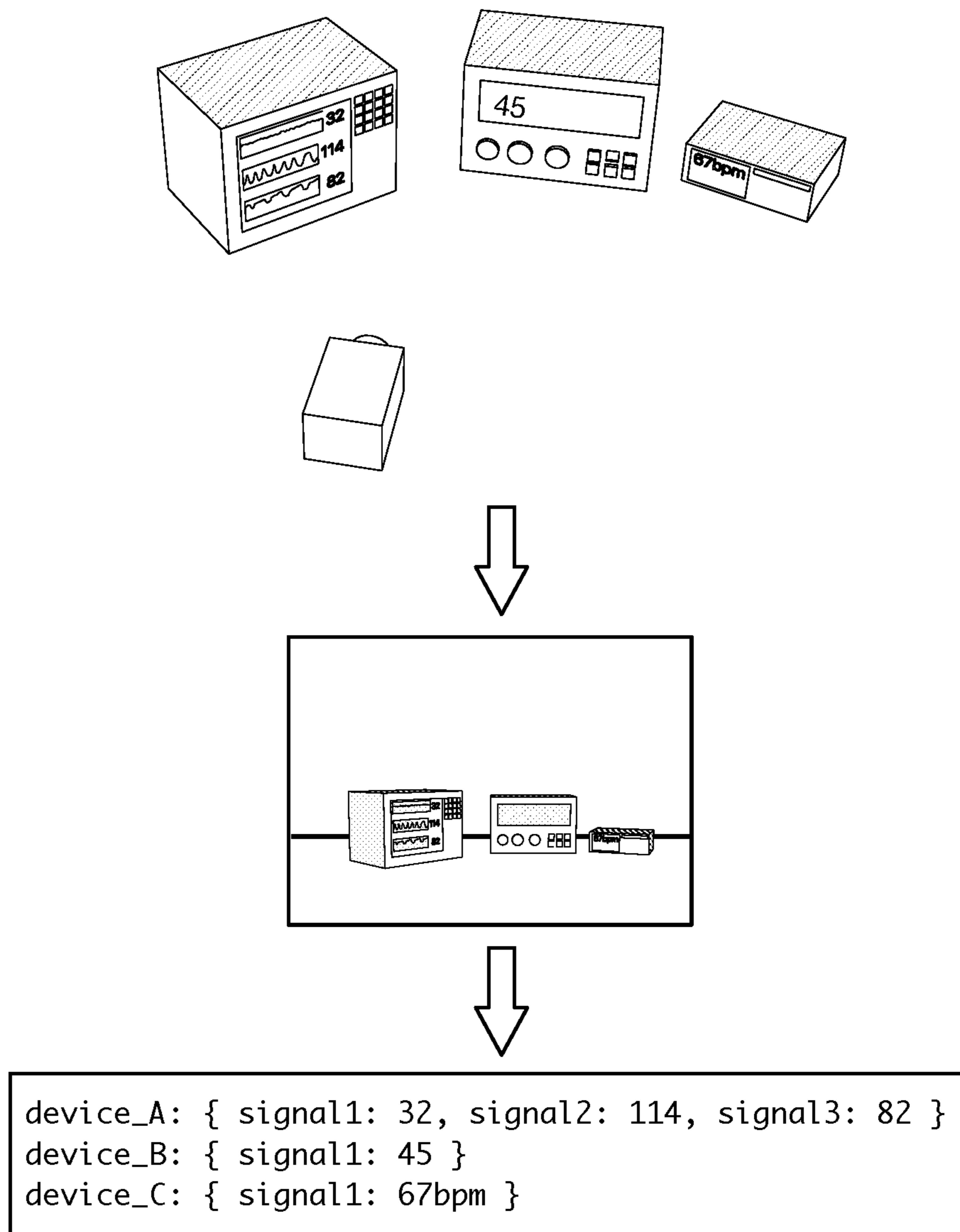


FIGURE 9



**FIGURE 10**



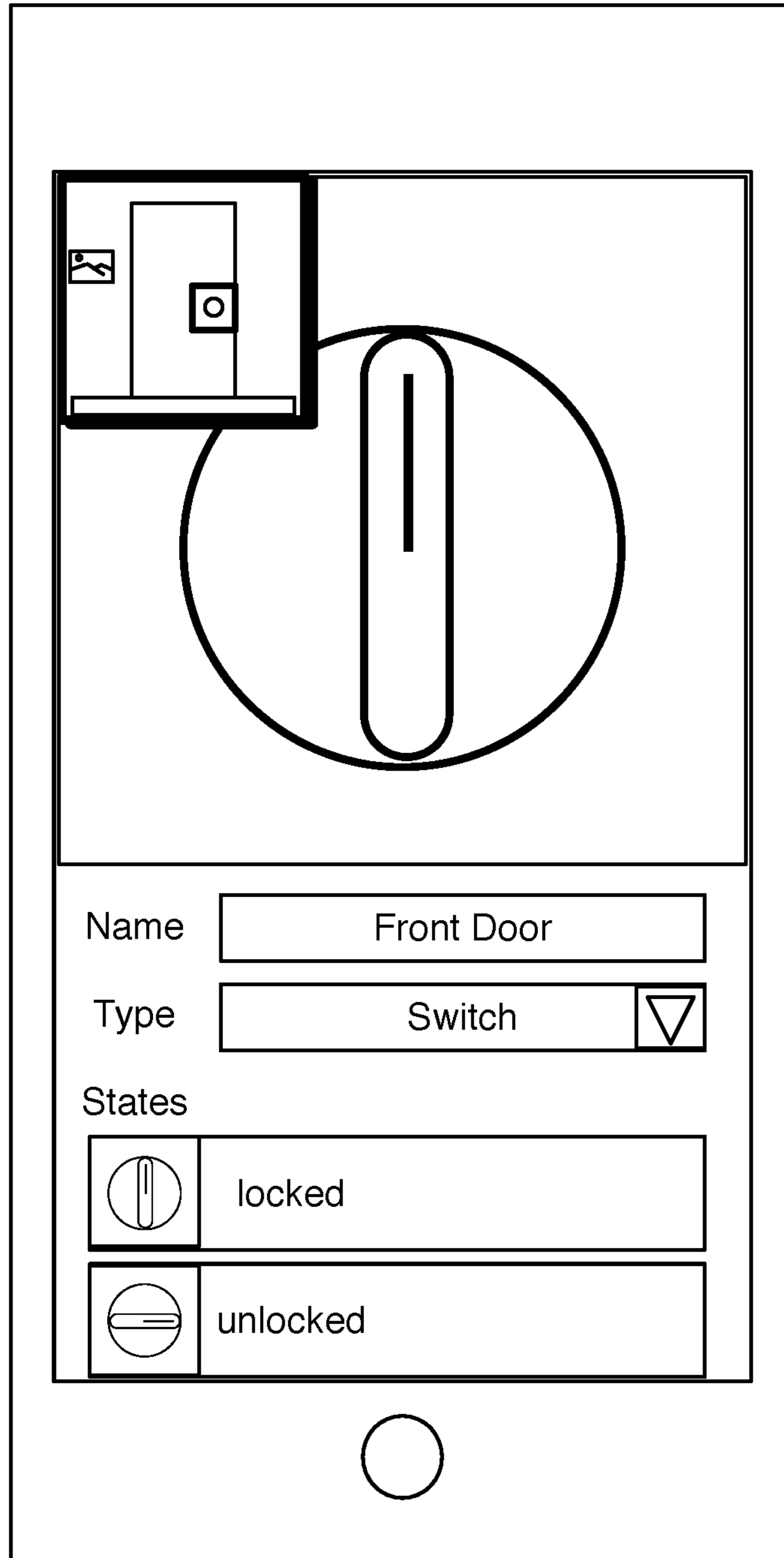
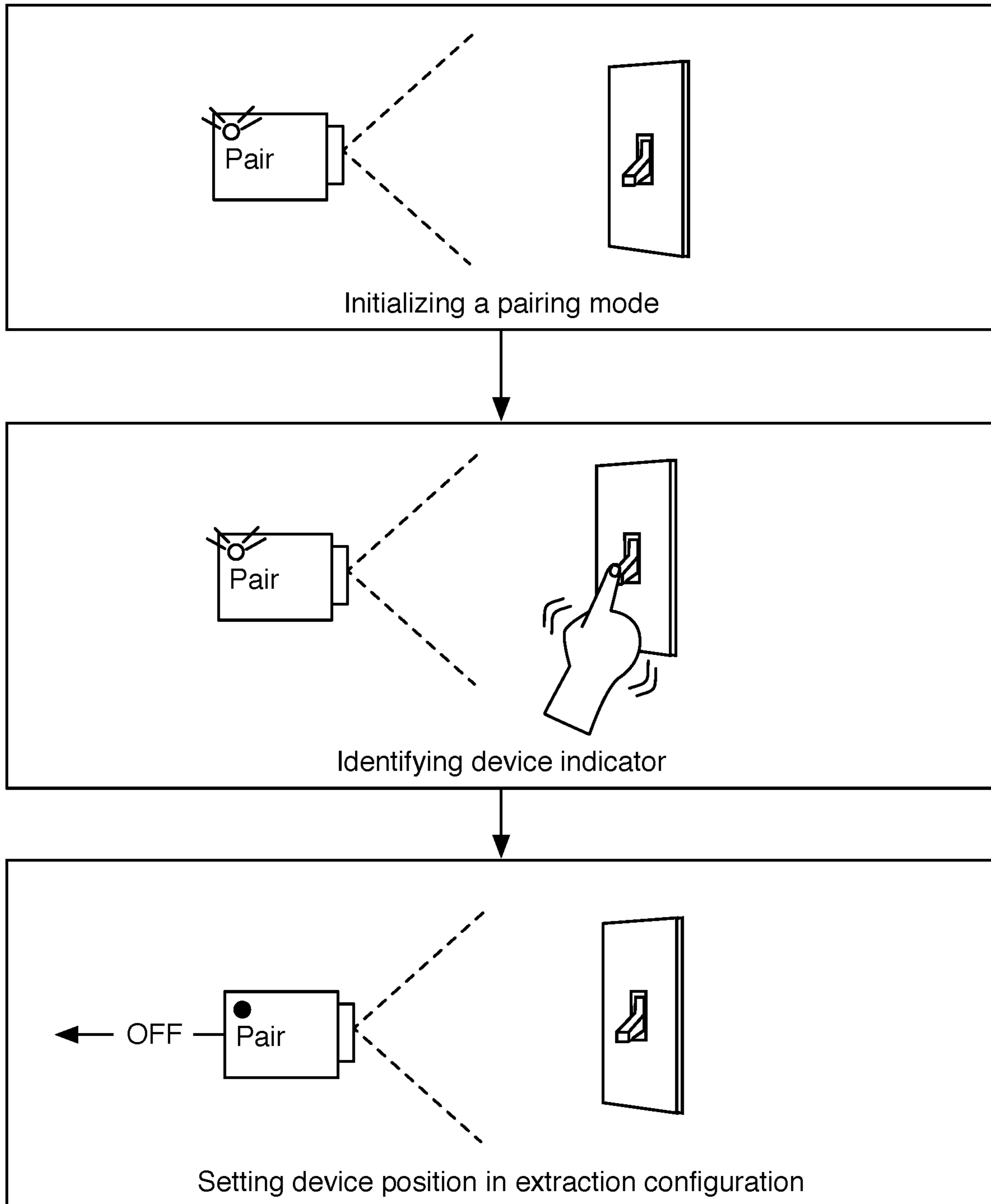
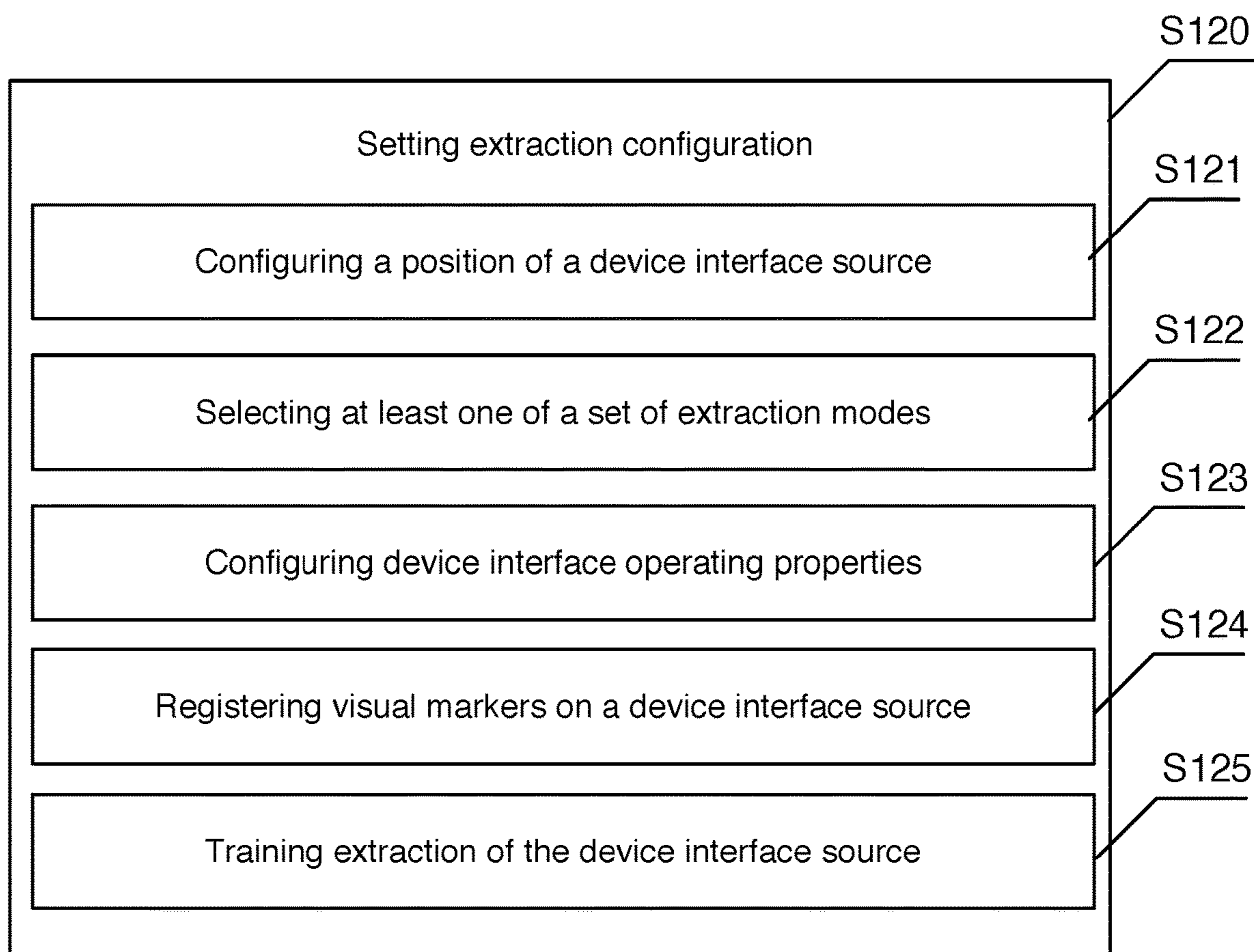


FIGURE 11



**FIGURE 12**



**FIGURE 13**



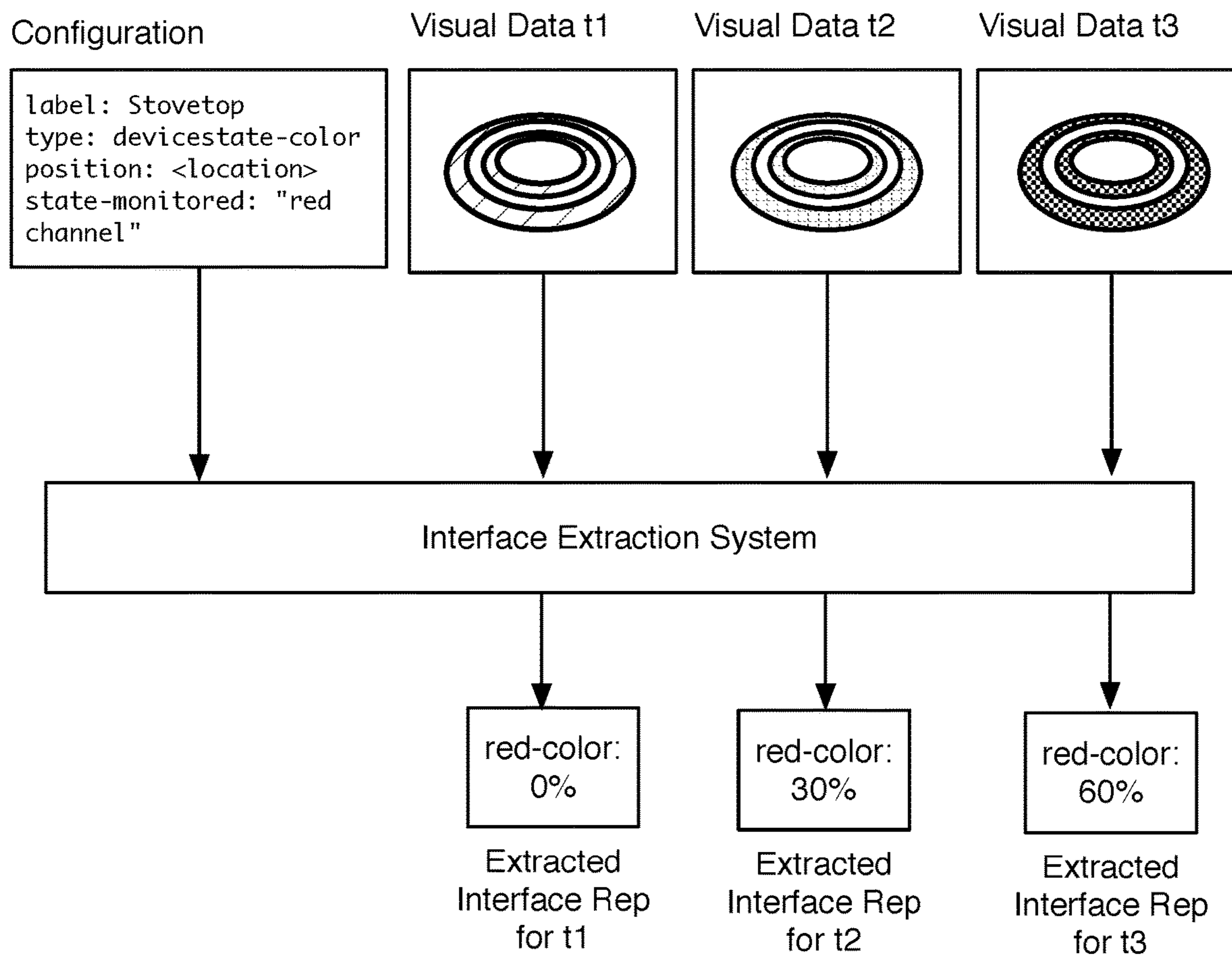


FIGURE 14

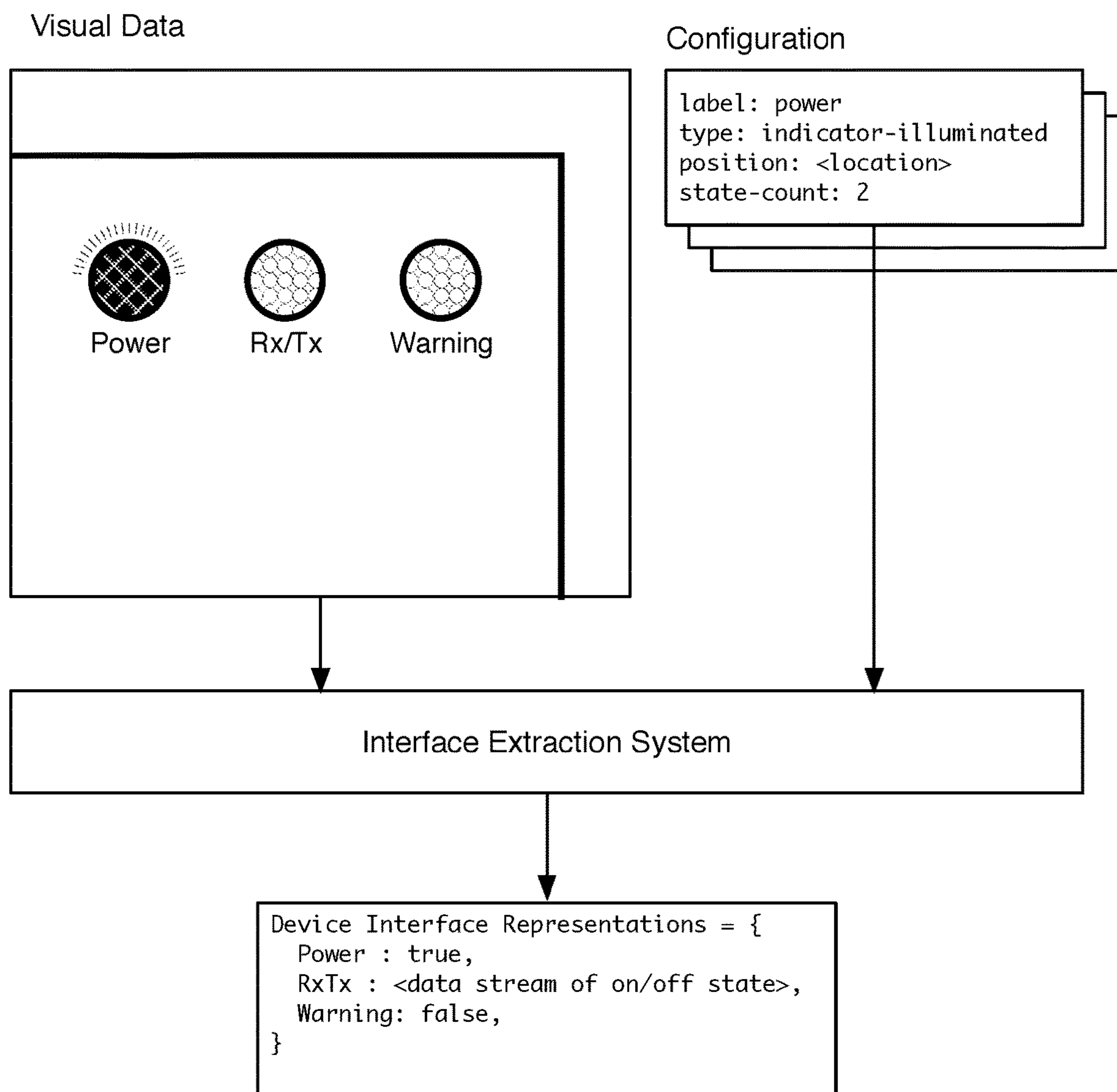


FIGURE 15

Configuration

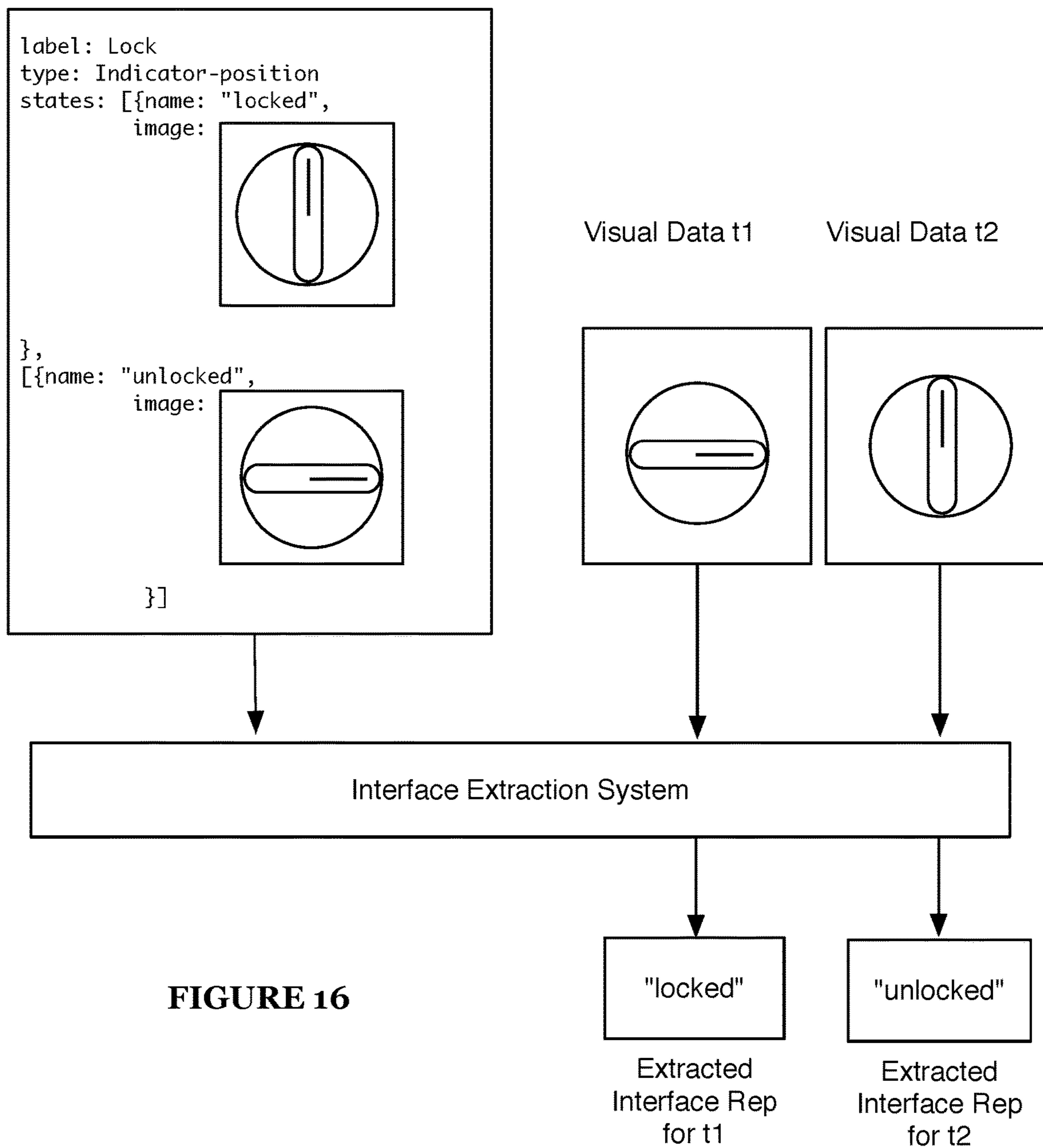
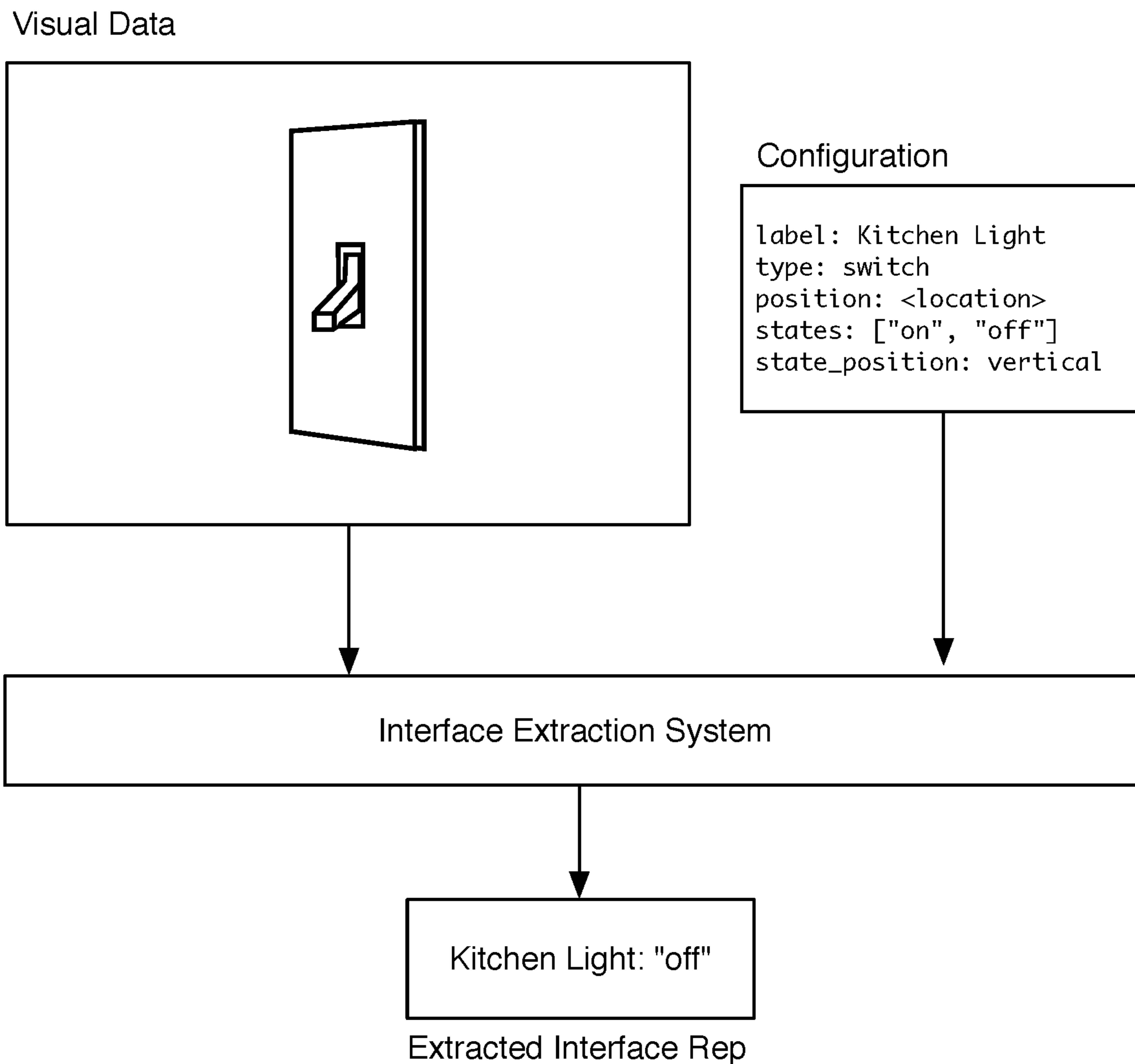


FIGURE 16





**FIGURE 17**

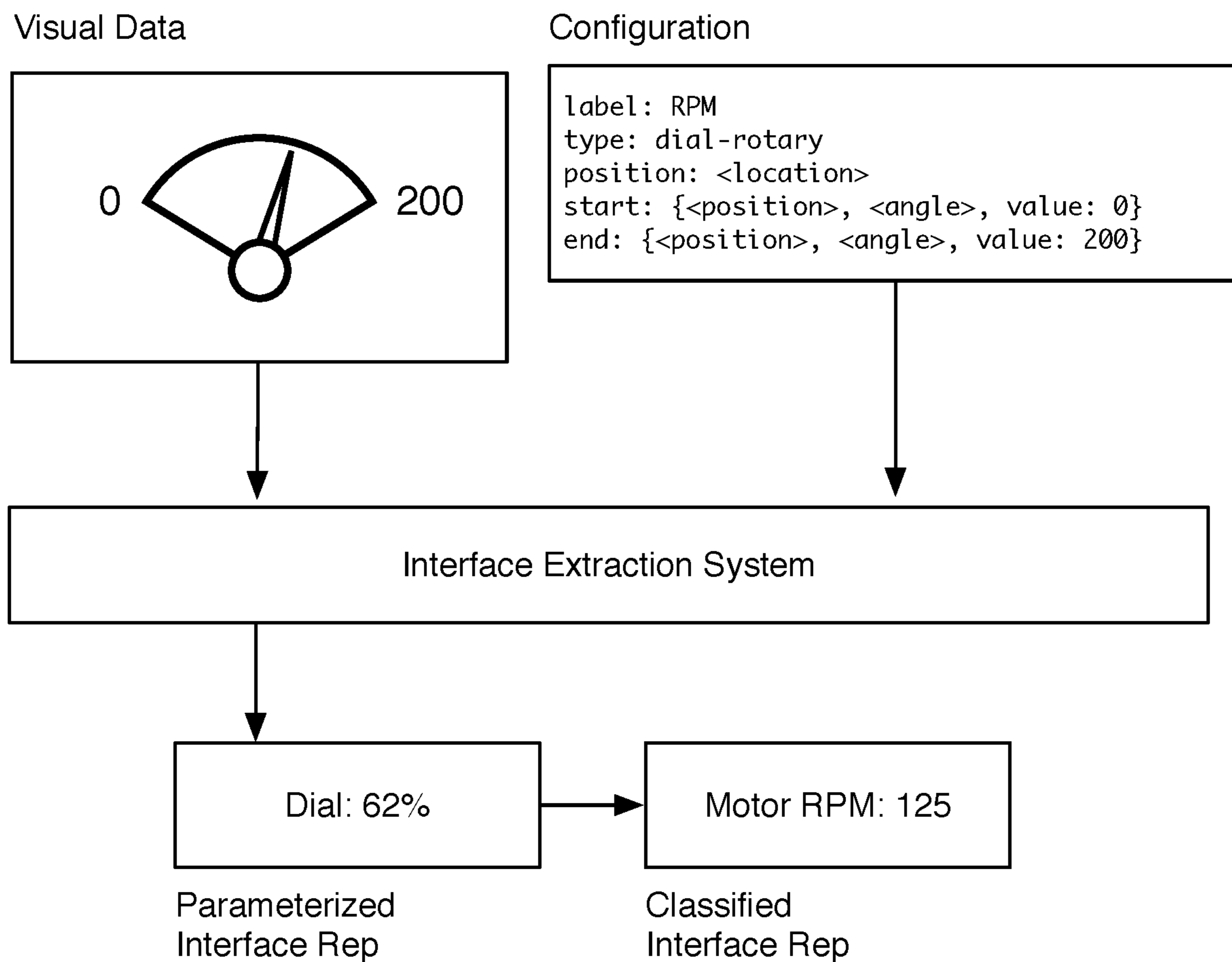
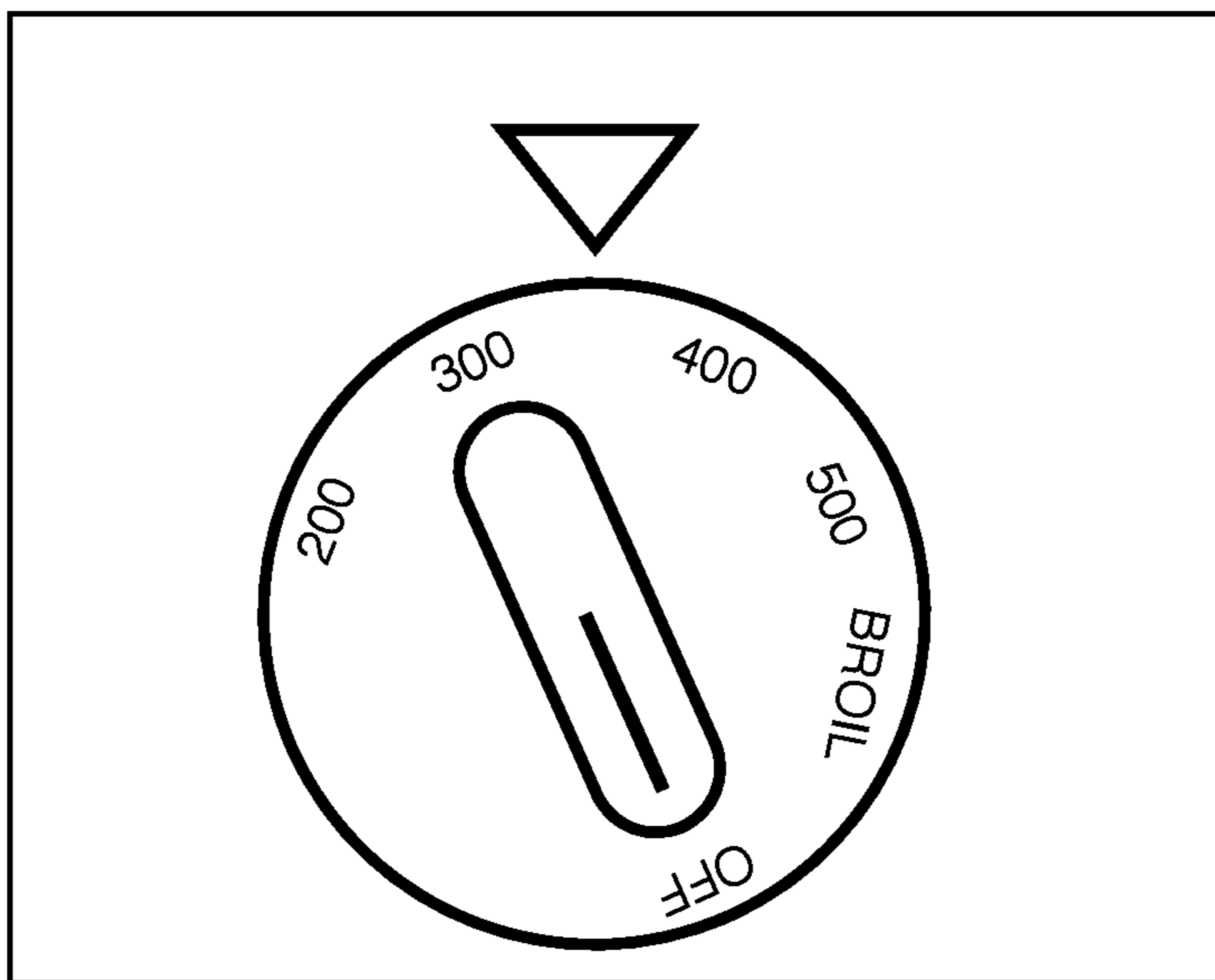


FIGURE 18

Visual Data



Configuration

```
label: Oven  
type: dial-rotary  
position: <location>  
min: 200  
max: 501  
indicator: <position>  
units: F
```

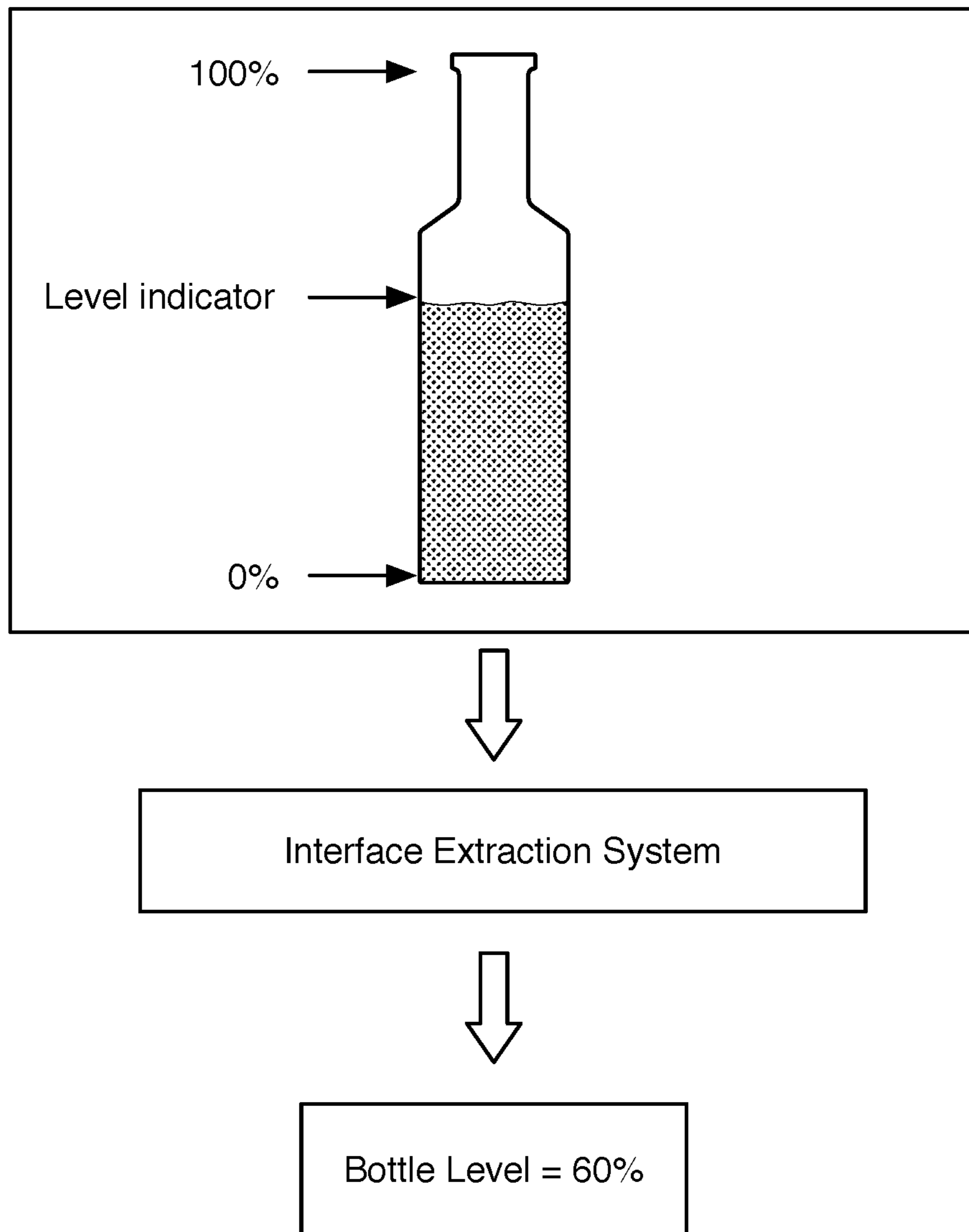
Interface Extraction System

Oven: 350° F

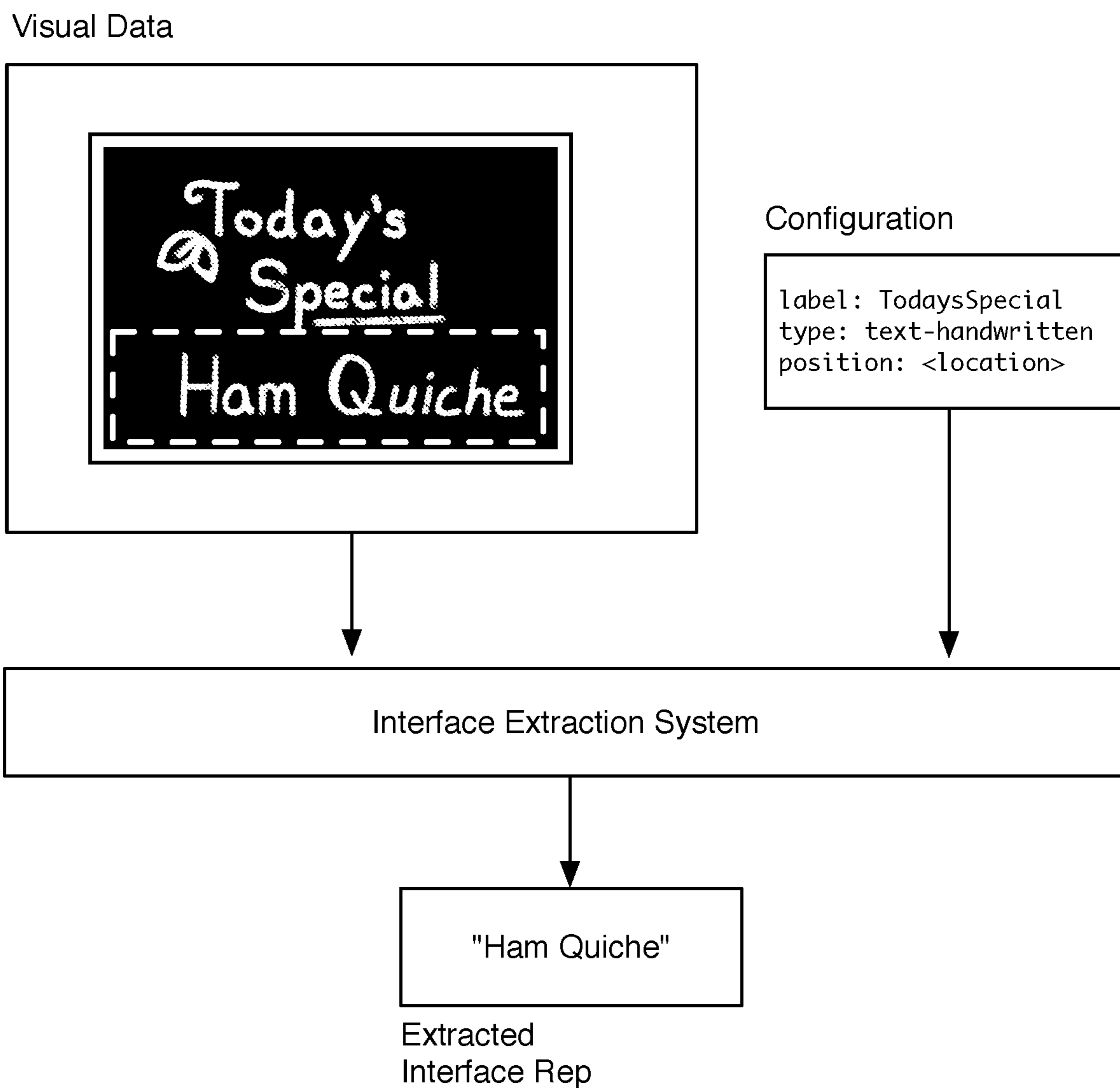
Extracted Interface Rep

**FIGURE 19**

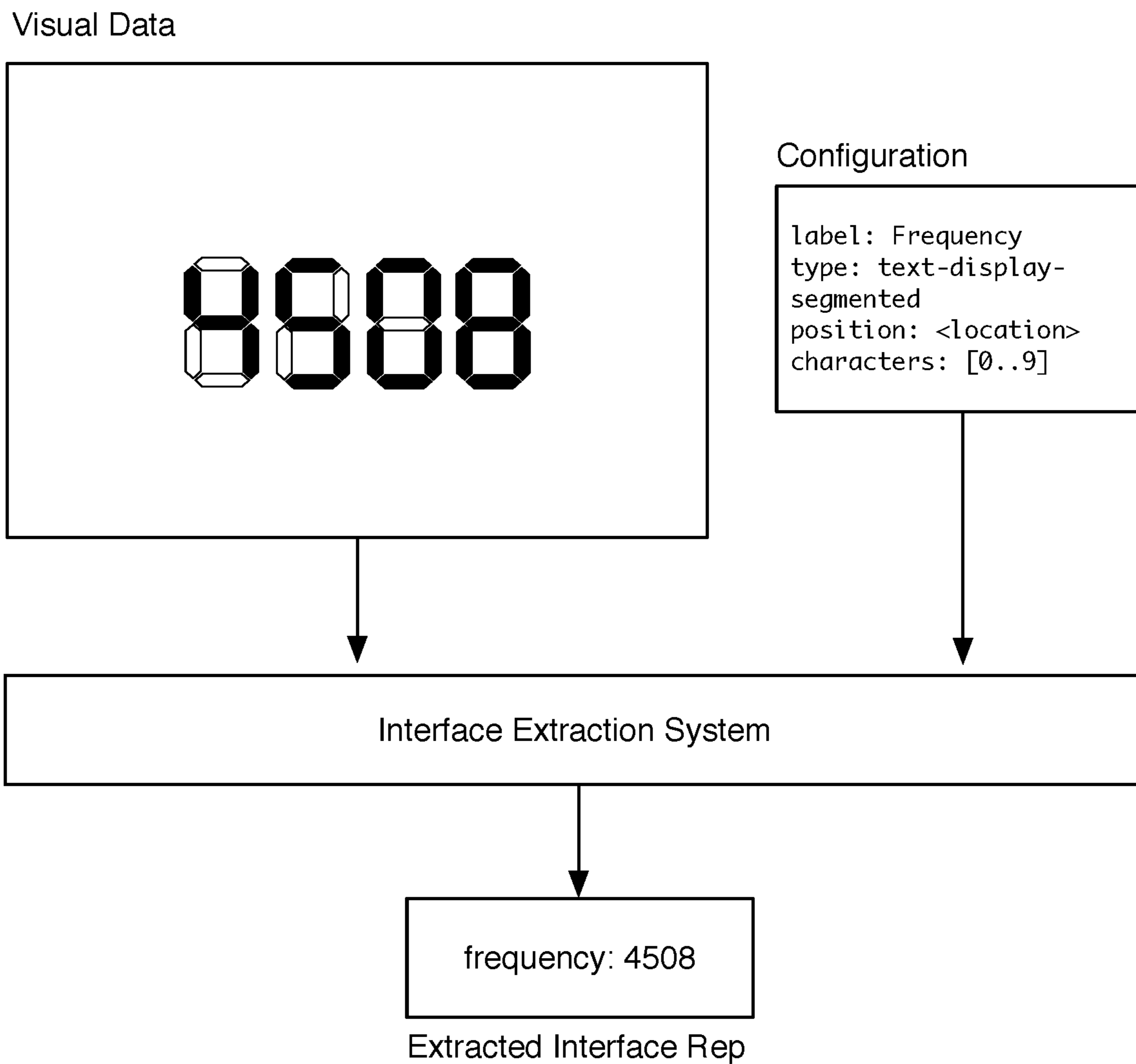




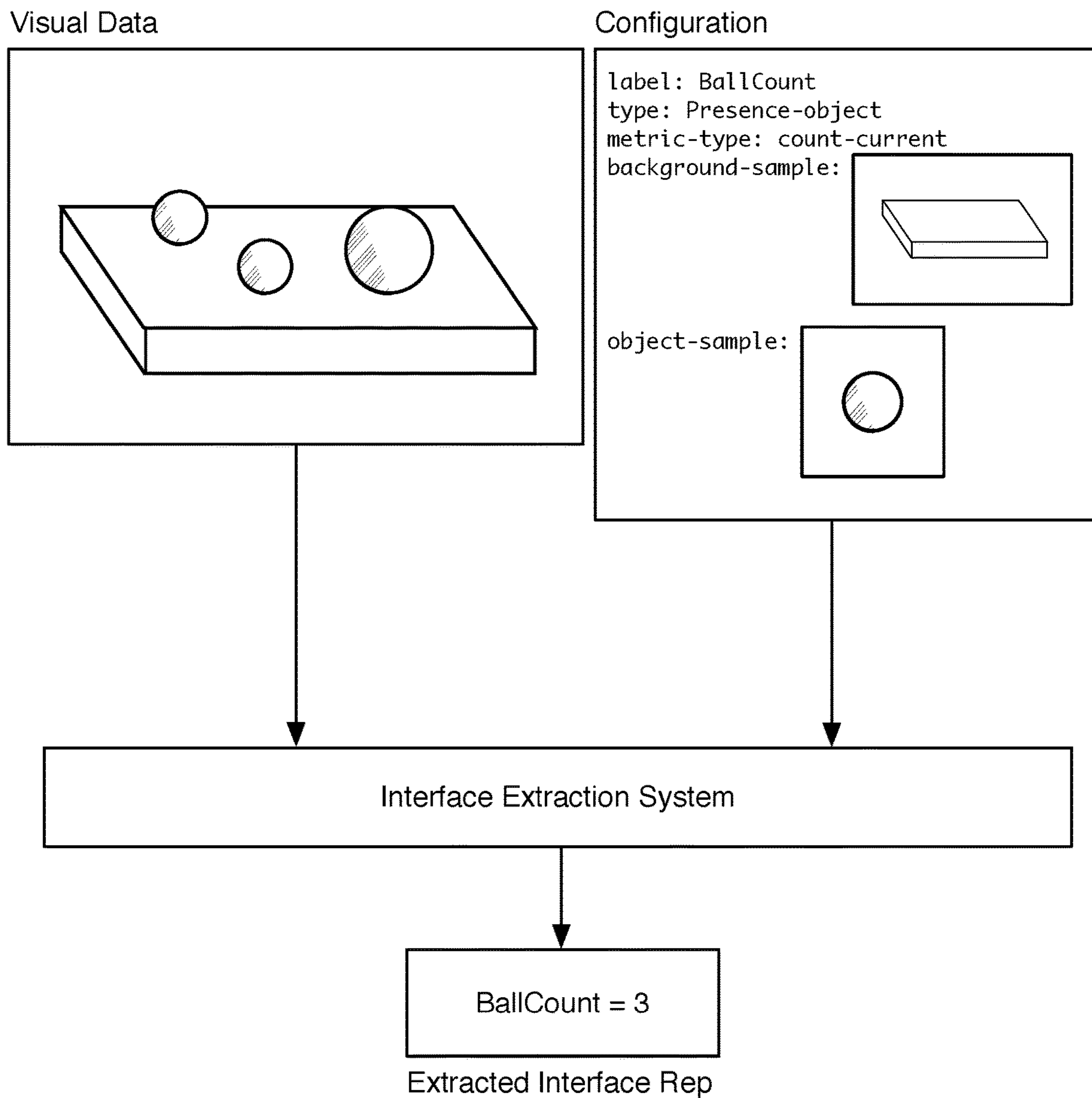
**FIGURE 20**



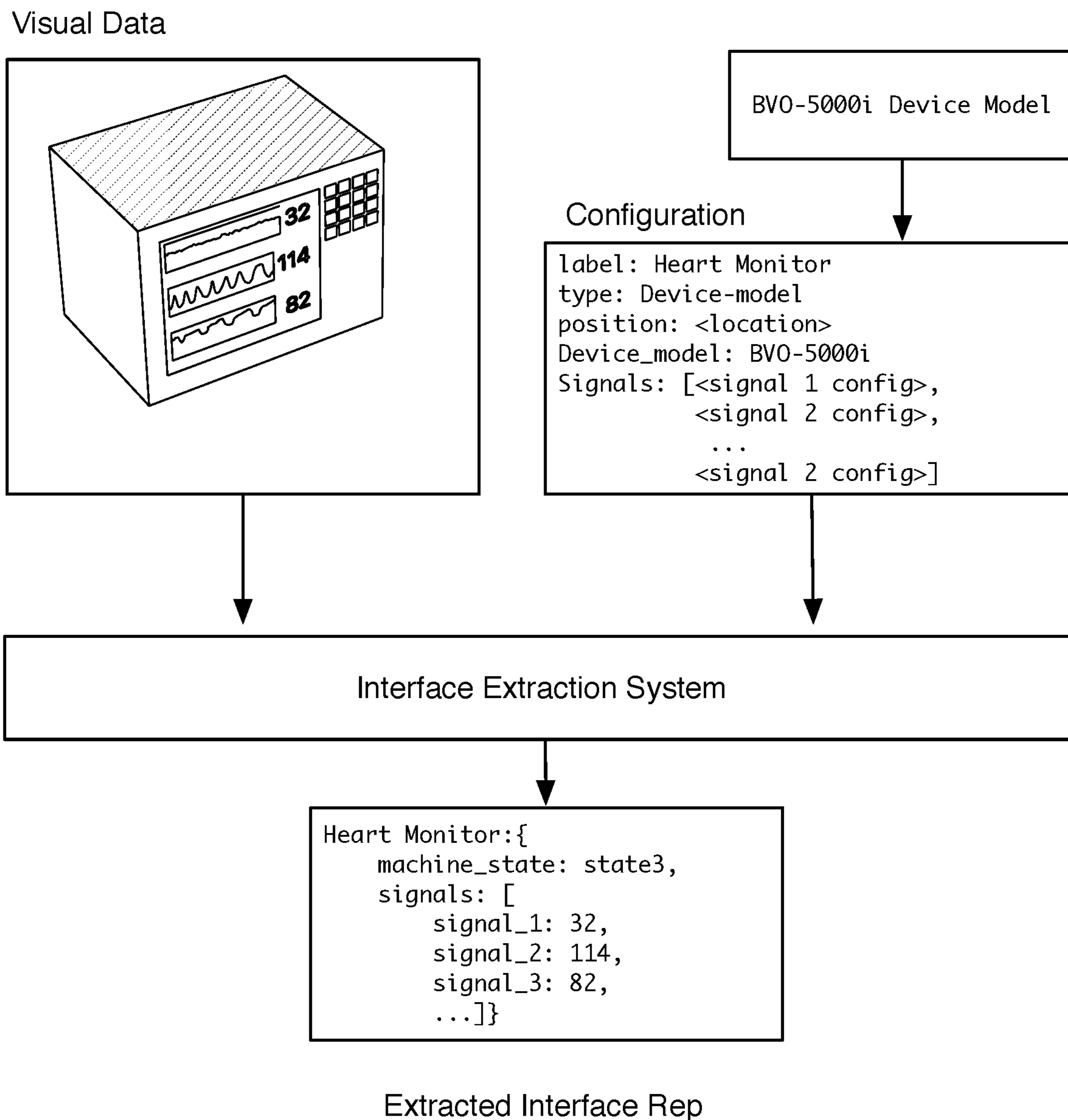
**FIGURE 21**



**FIGURE 22**

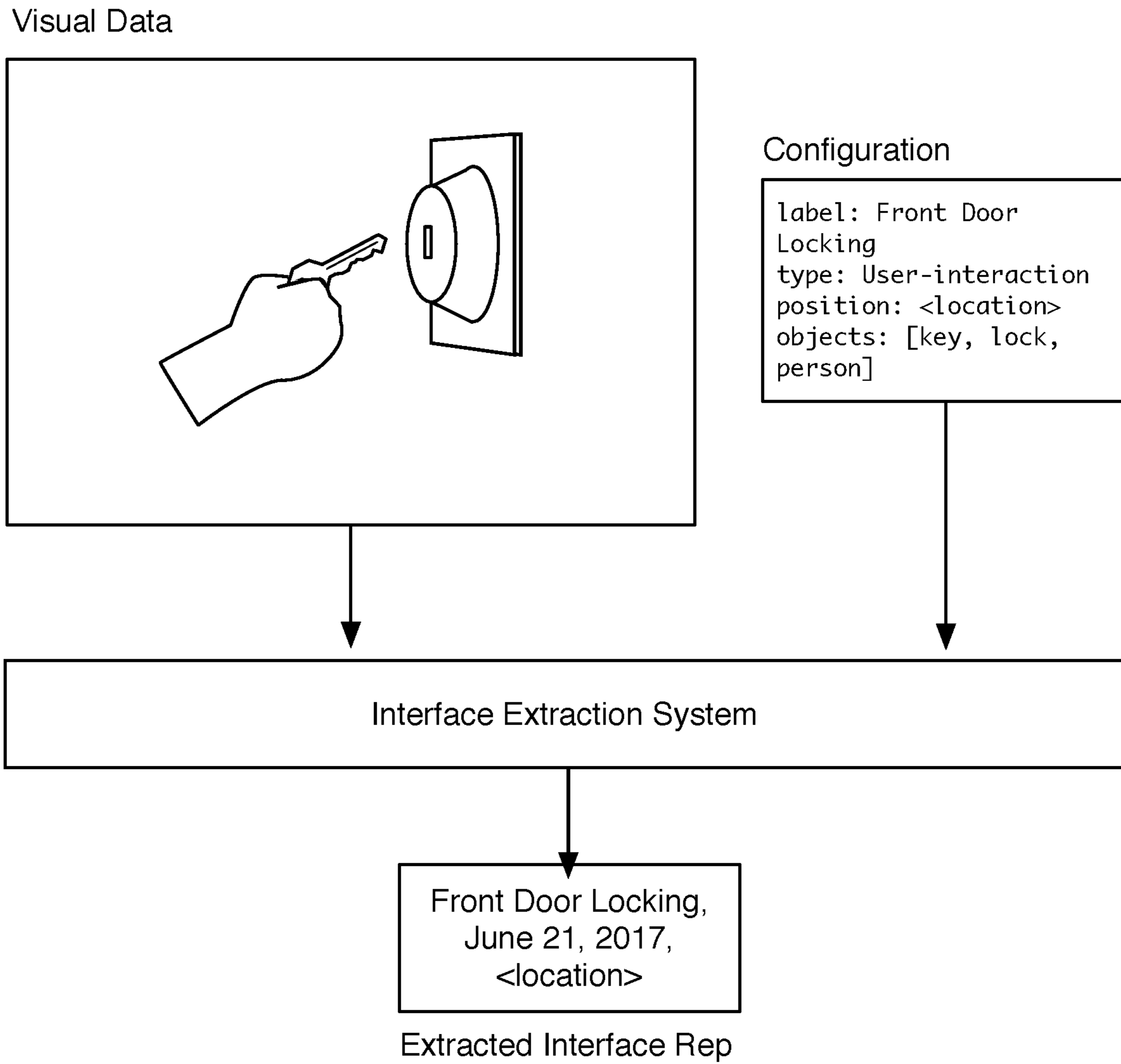


**FIGURE 23**

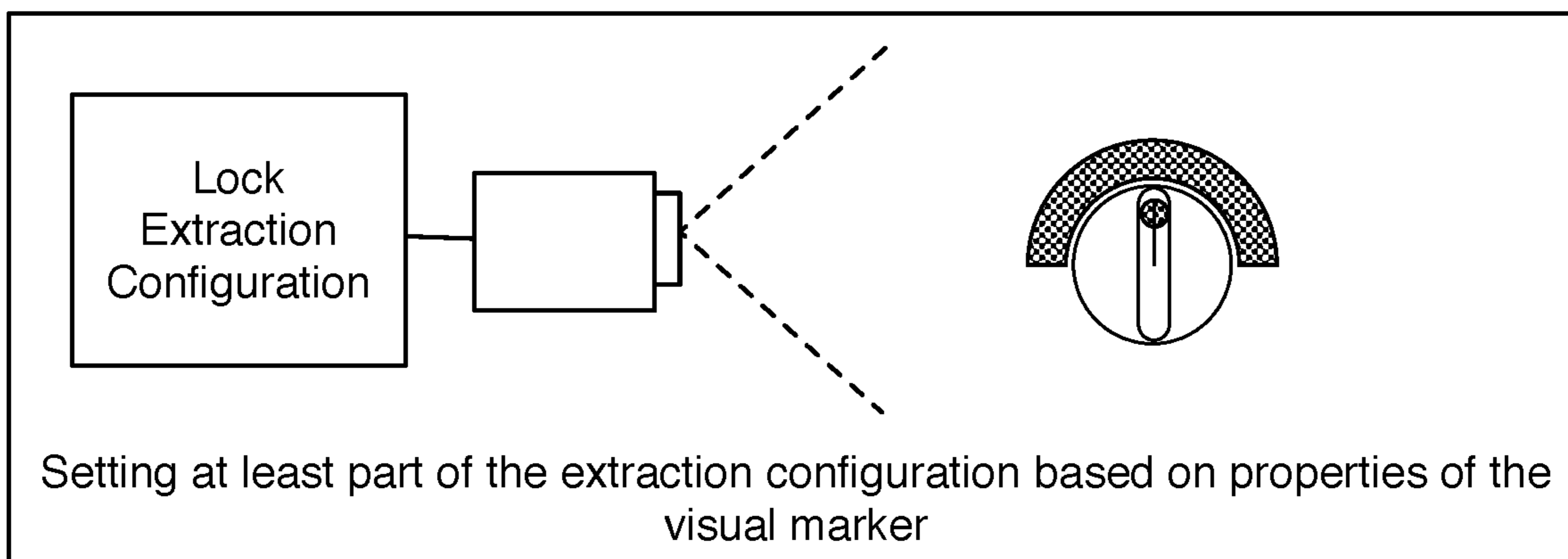
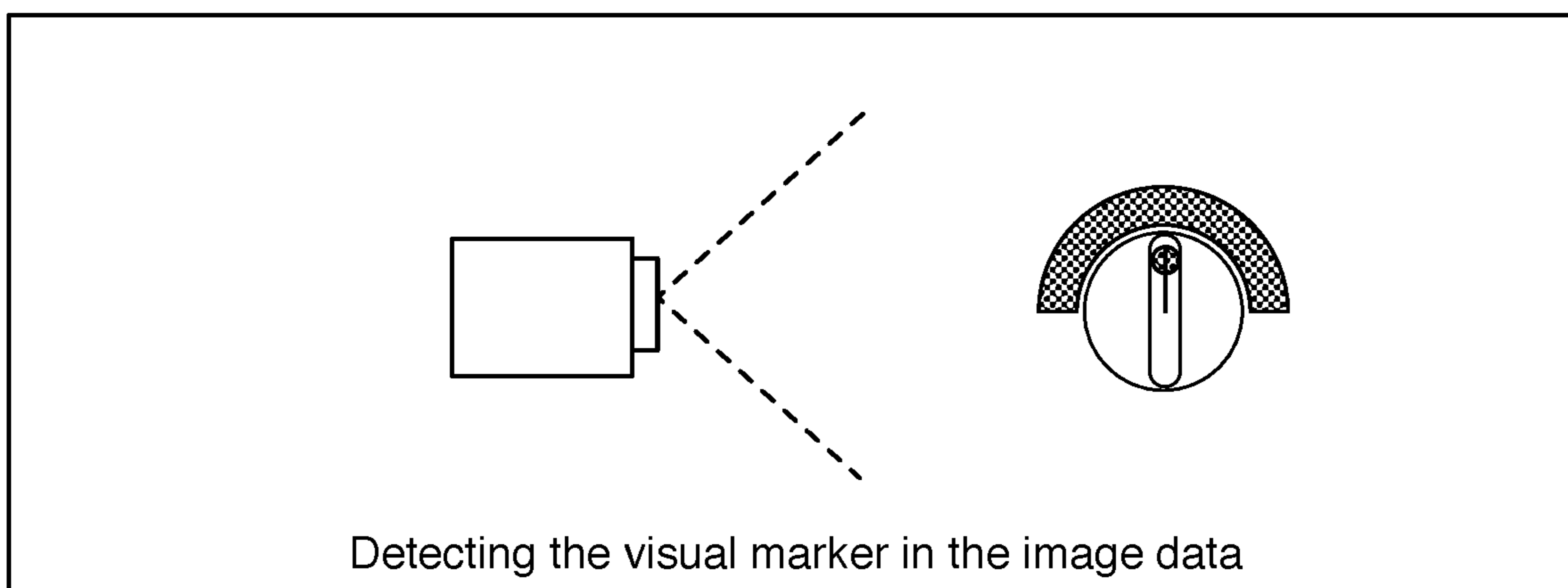
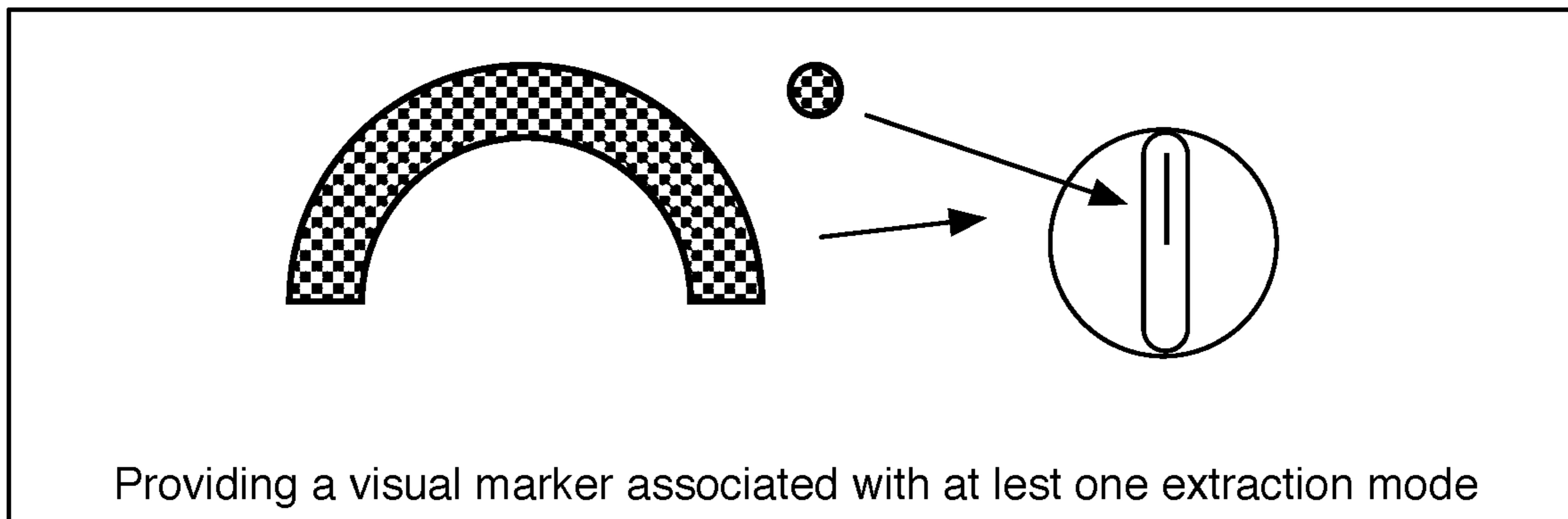


**FIGURE 24**





**FIGURE 25**



**FIGURE 26**

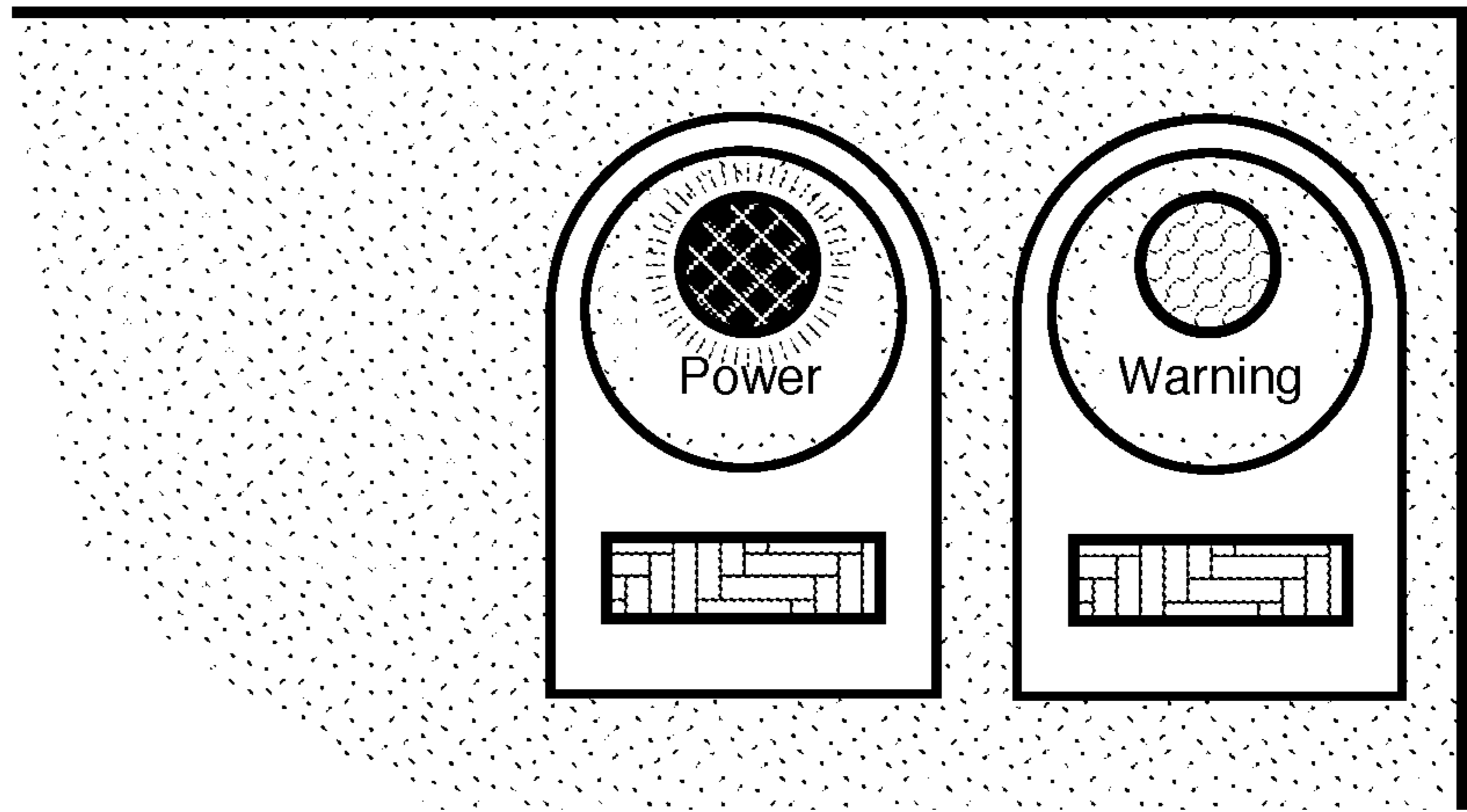


FIGURE 27A

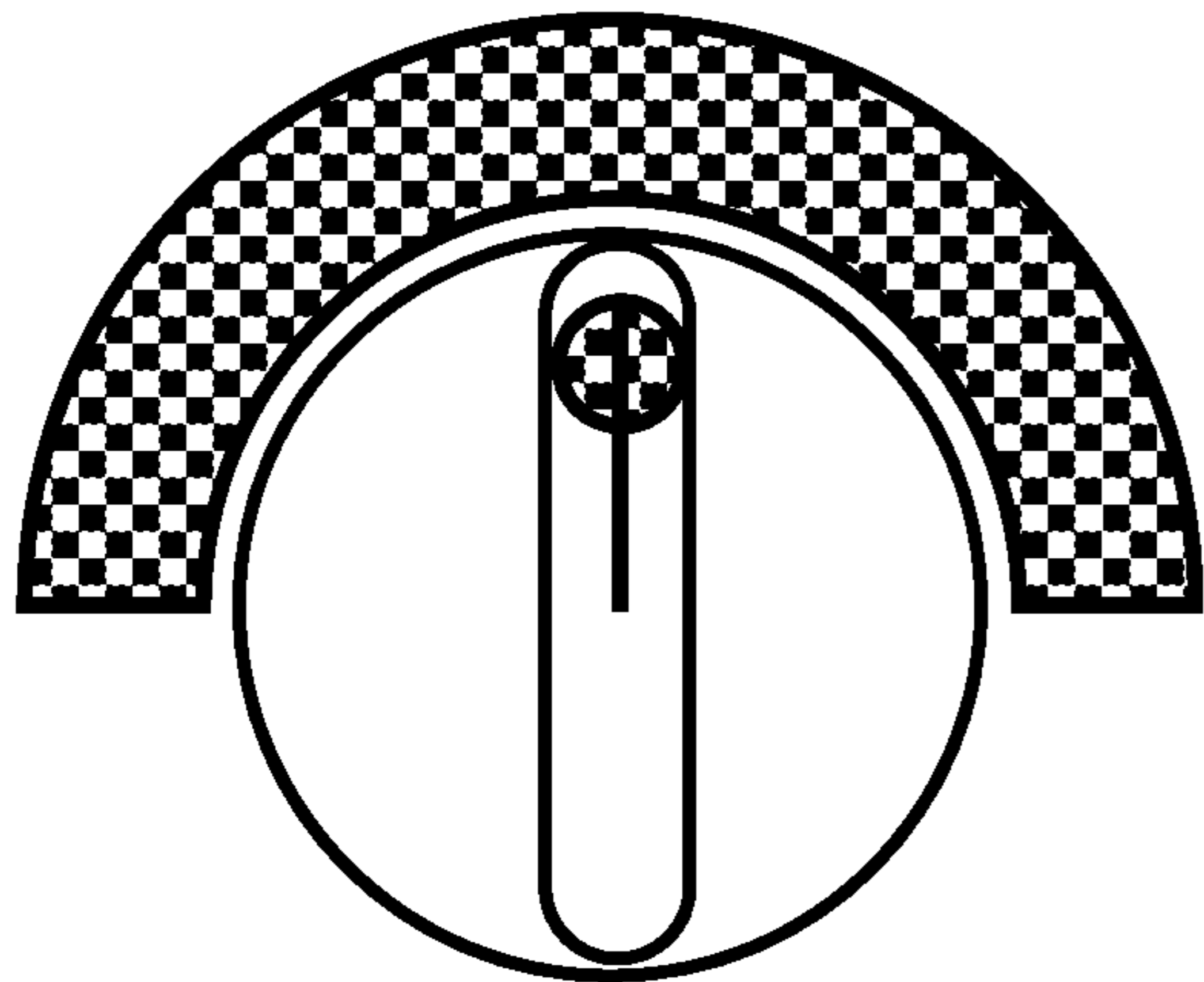


FIGURE 27B

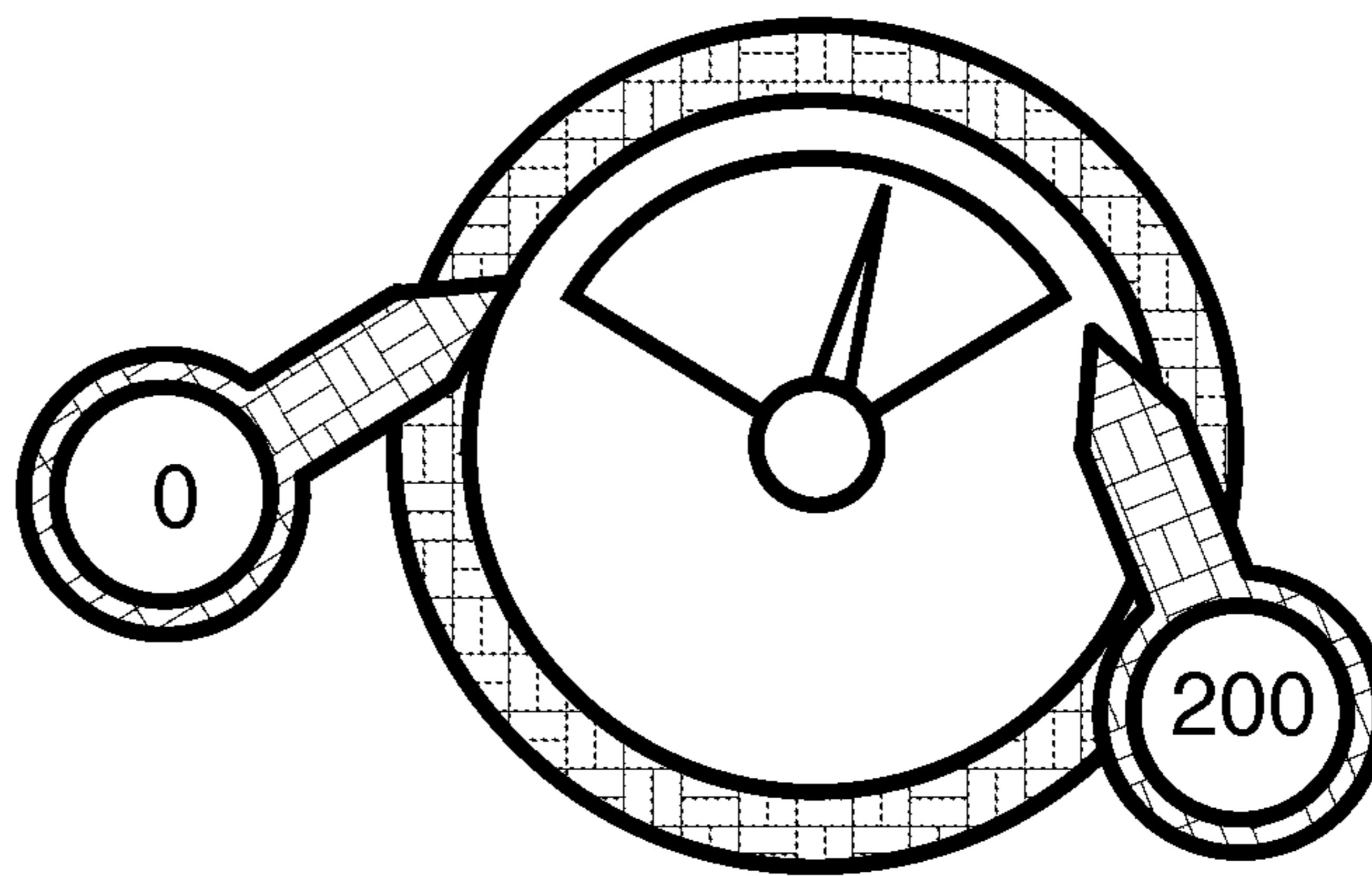
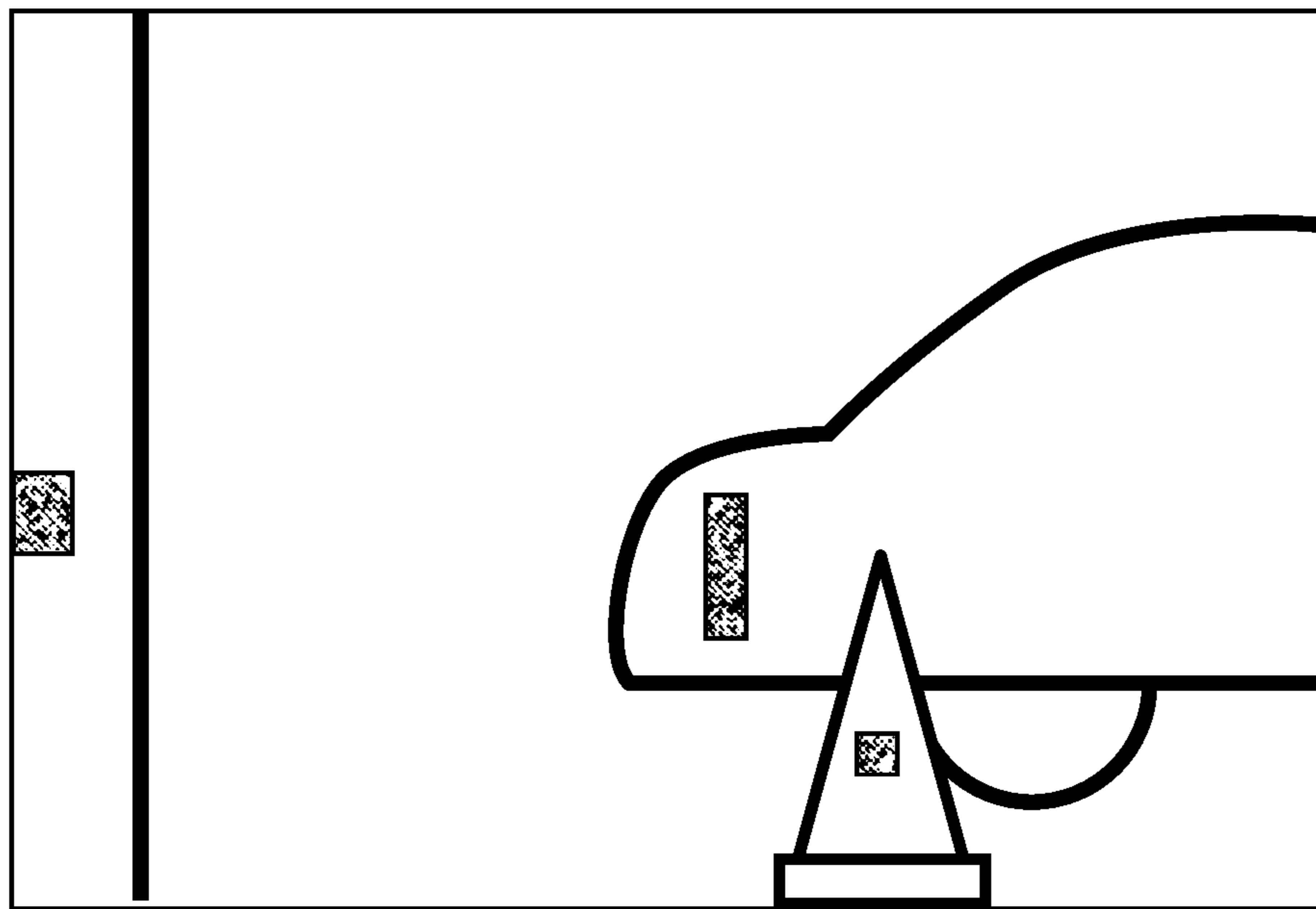
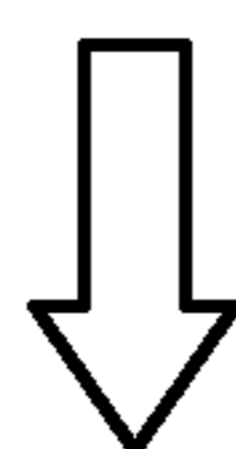
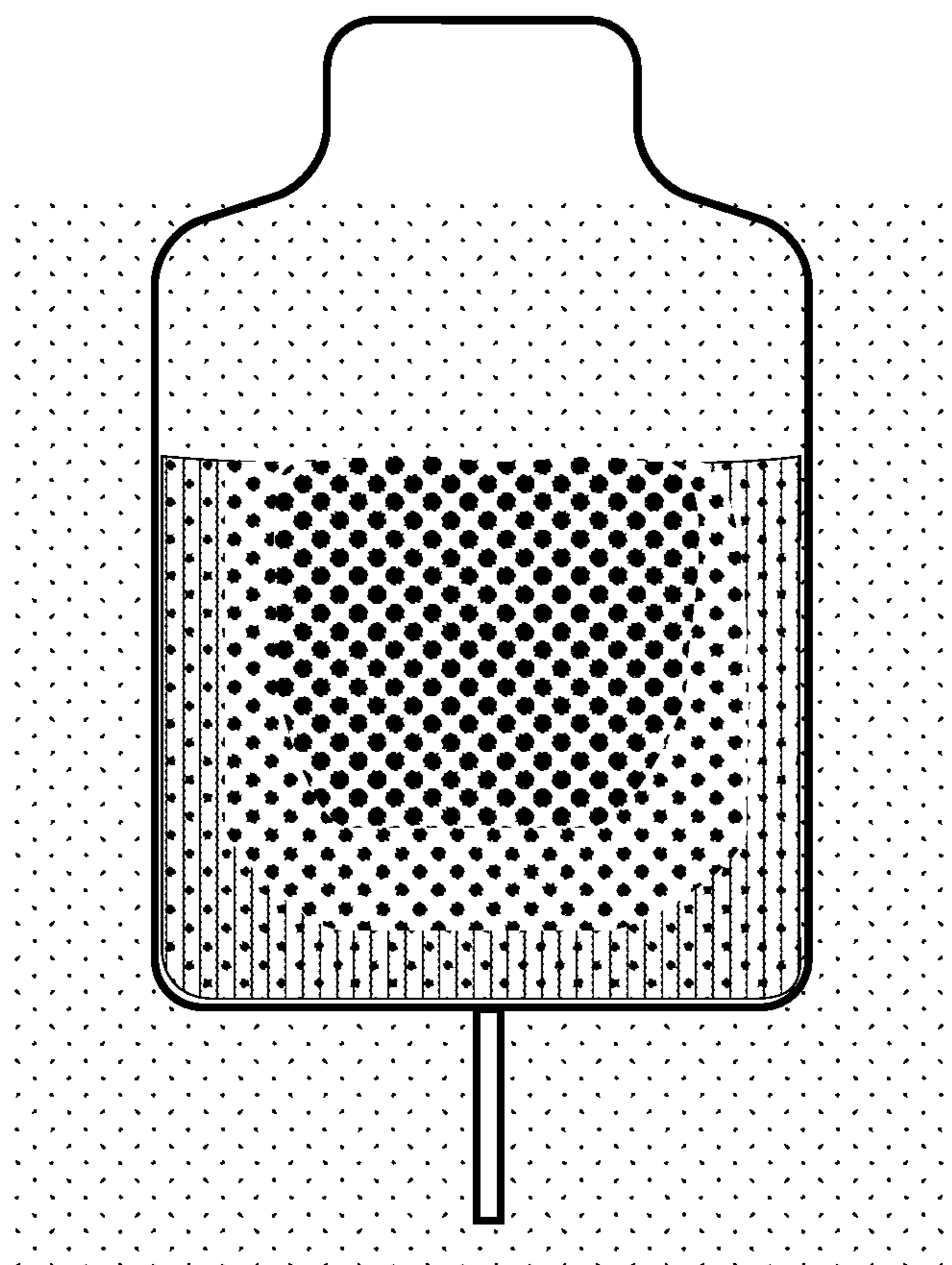


FIGURE 27C



Car Position = 5ft

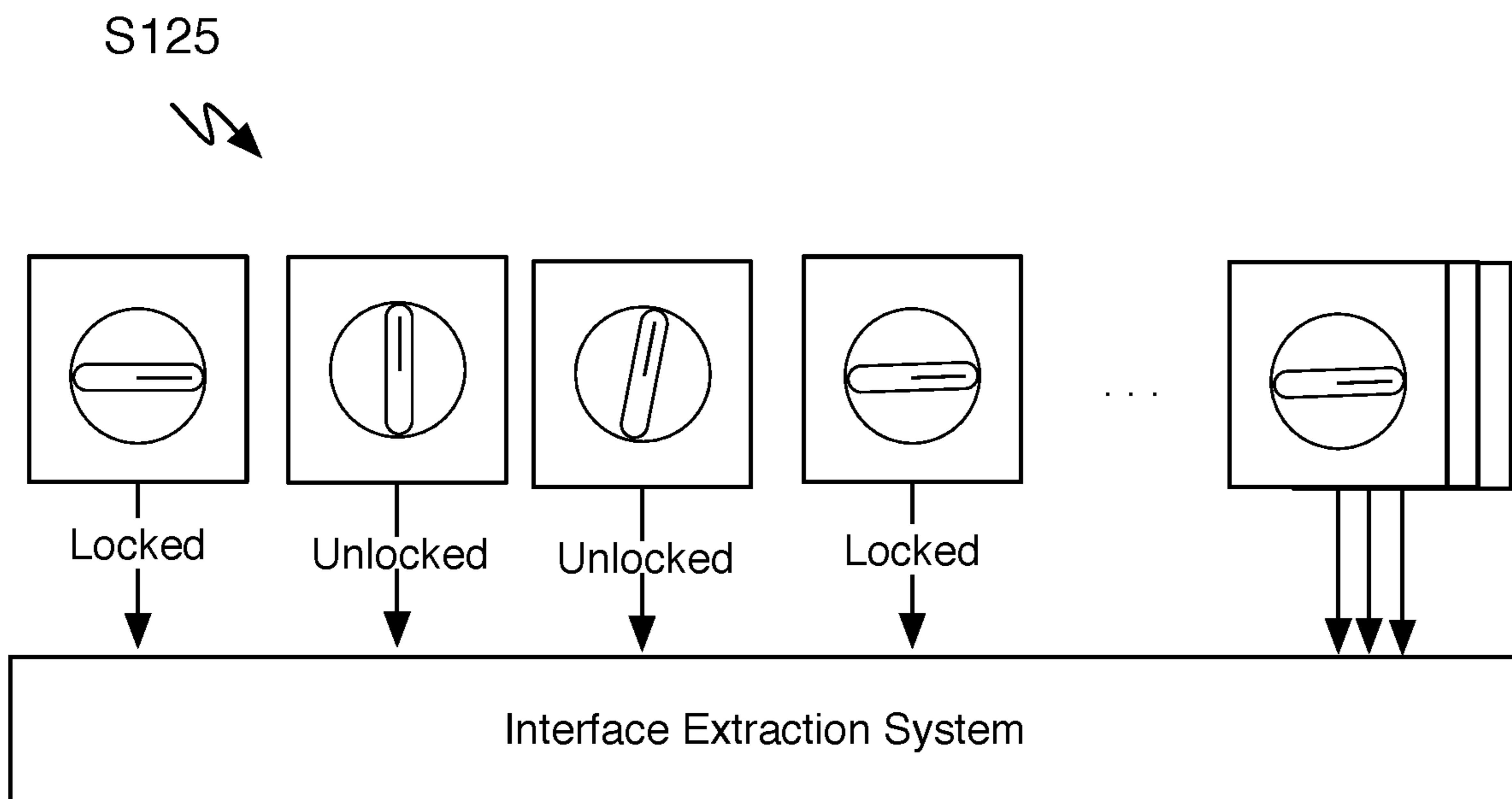
**FIGURE 28**



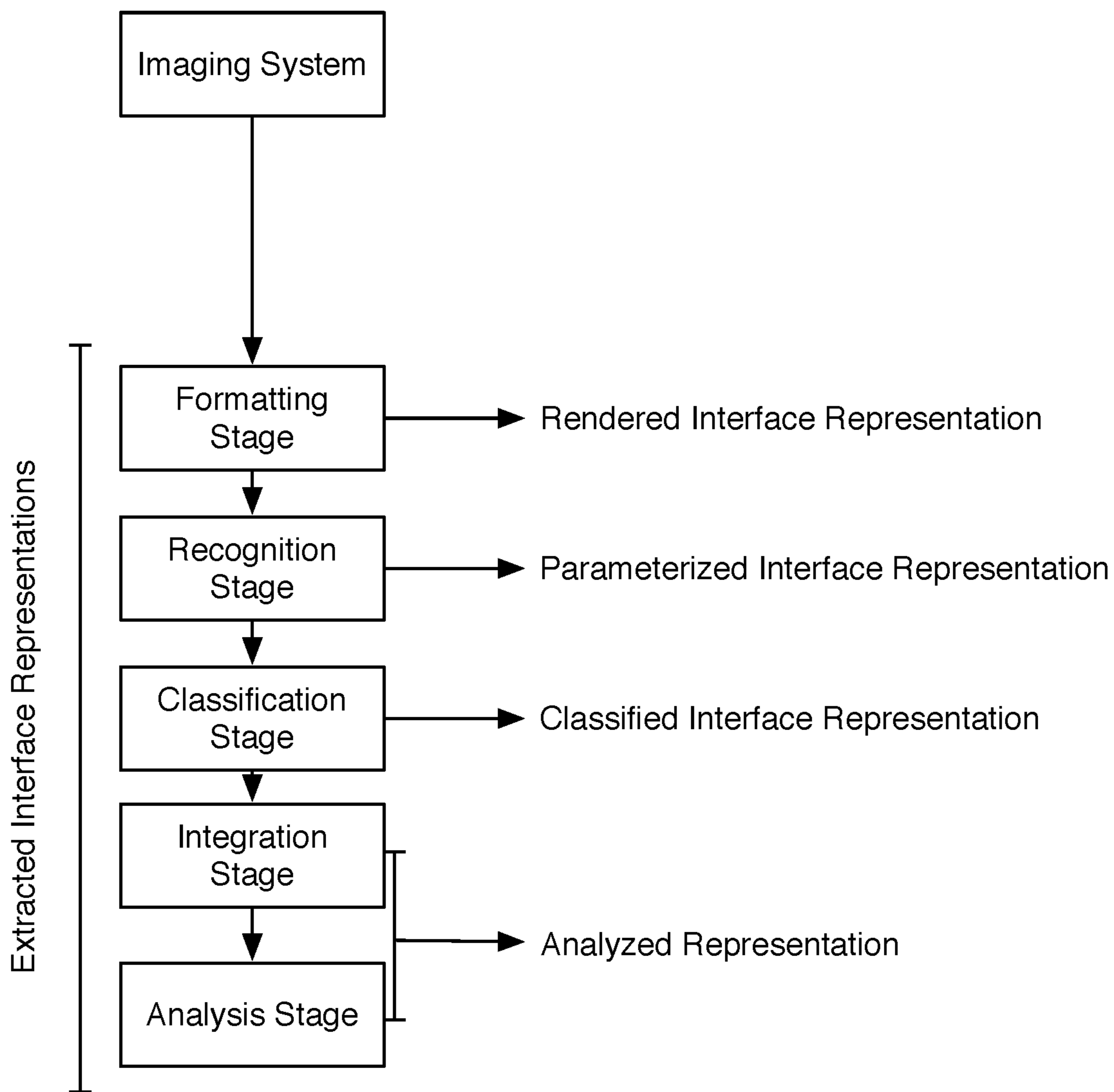
IV Level = 70%

**FIGURE 29**

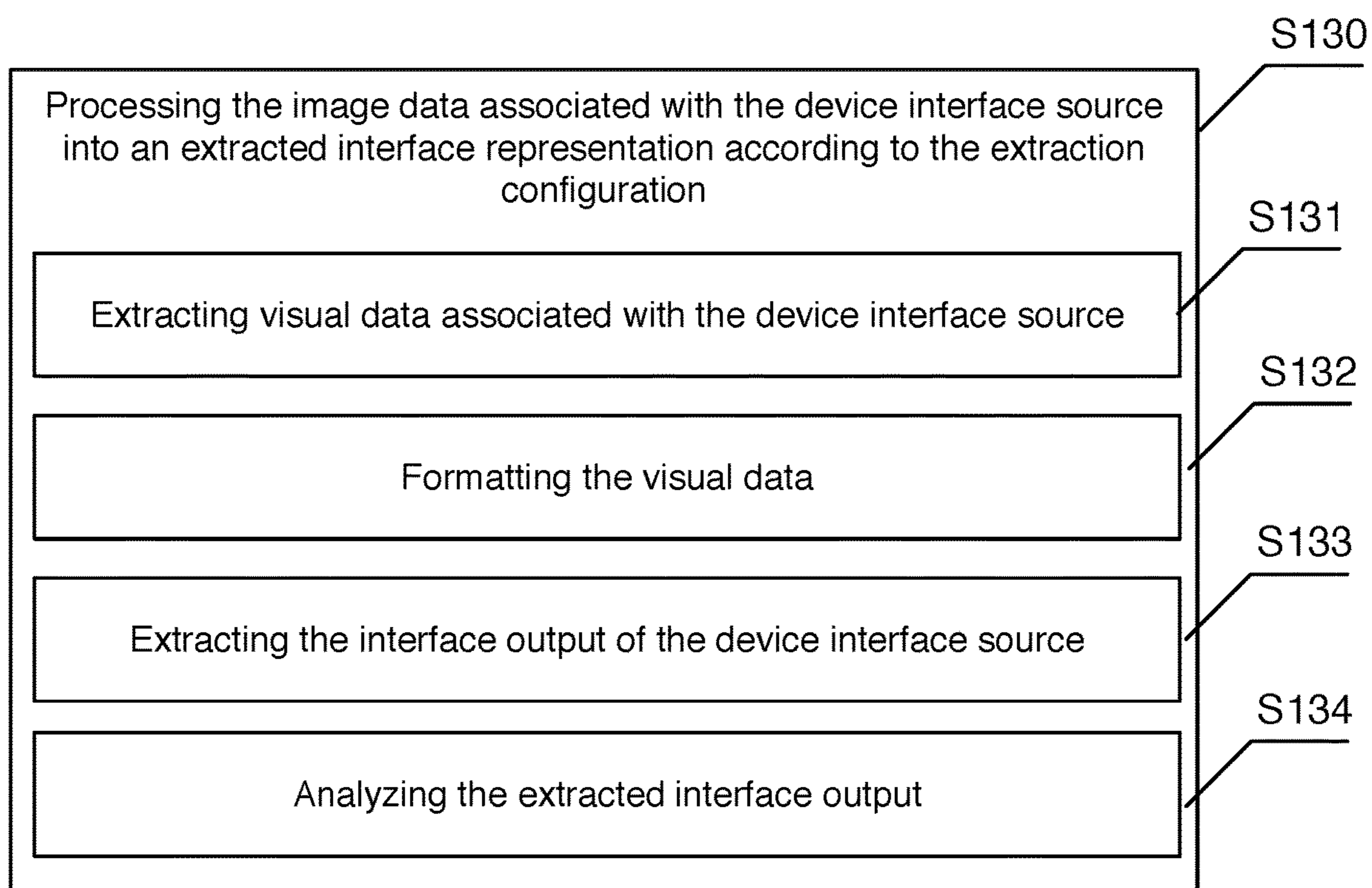




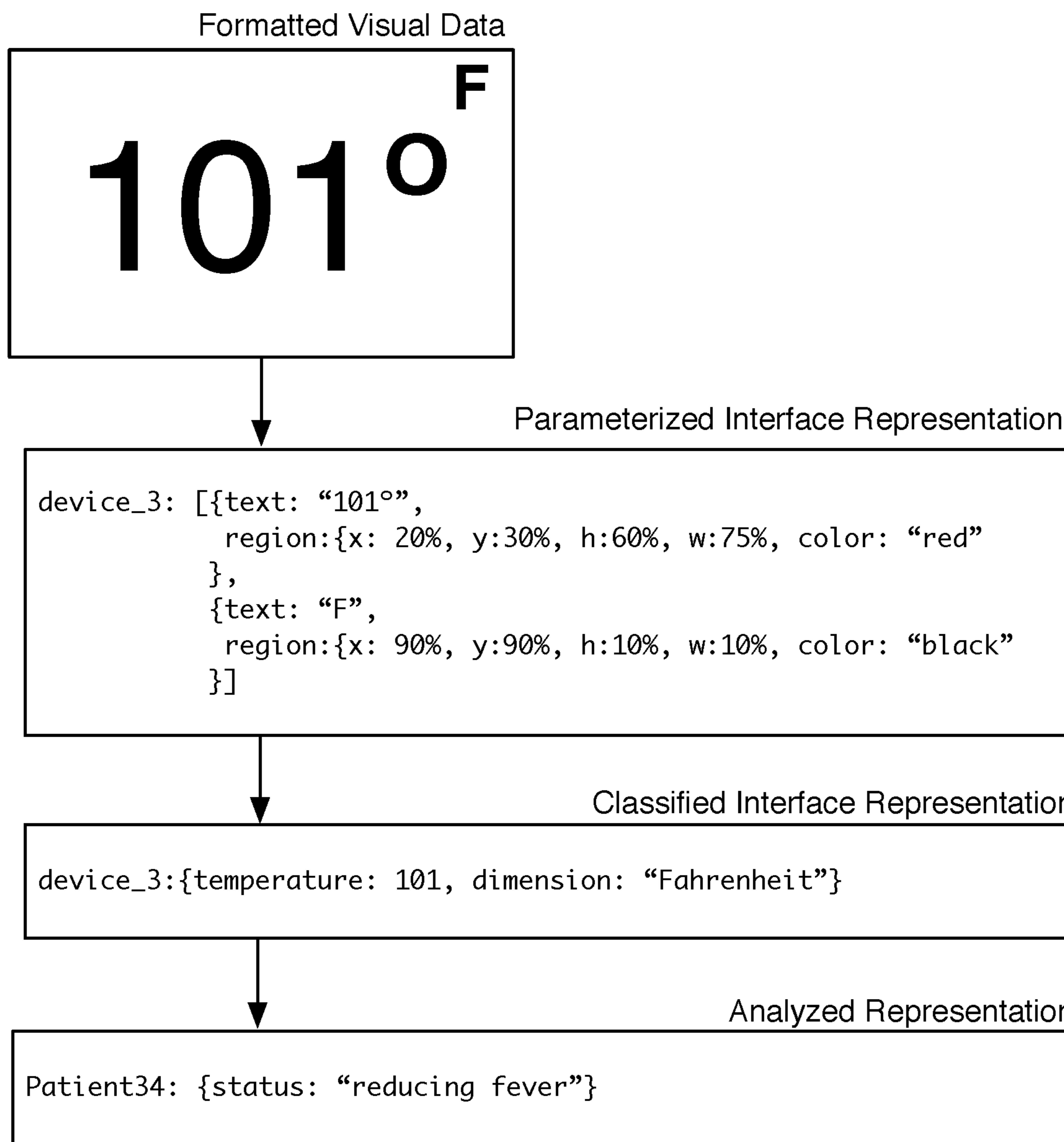
**FIGURE 30**



**FIGURE 31**



**FIGURE 32**



**FIGURE 33**

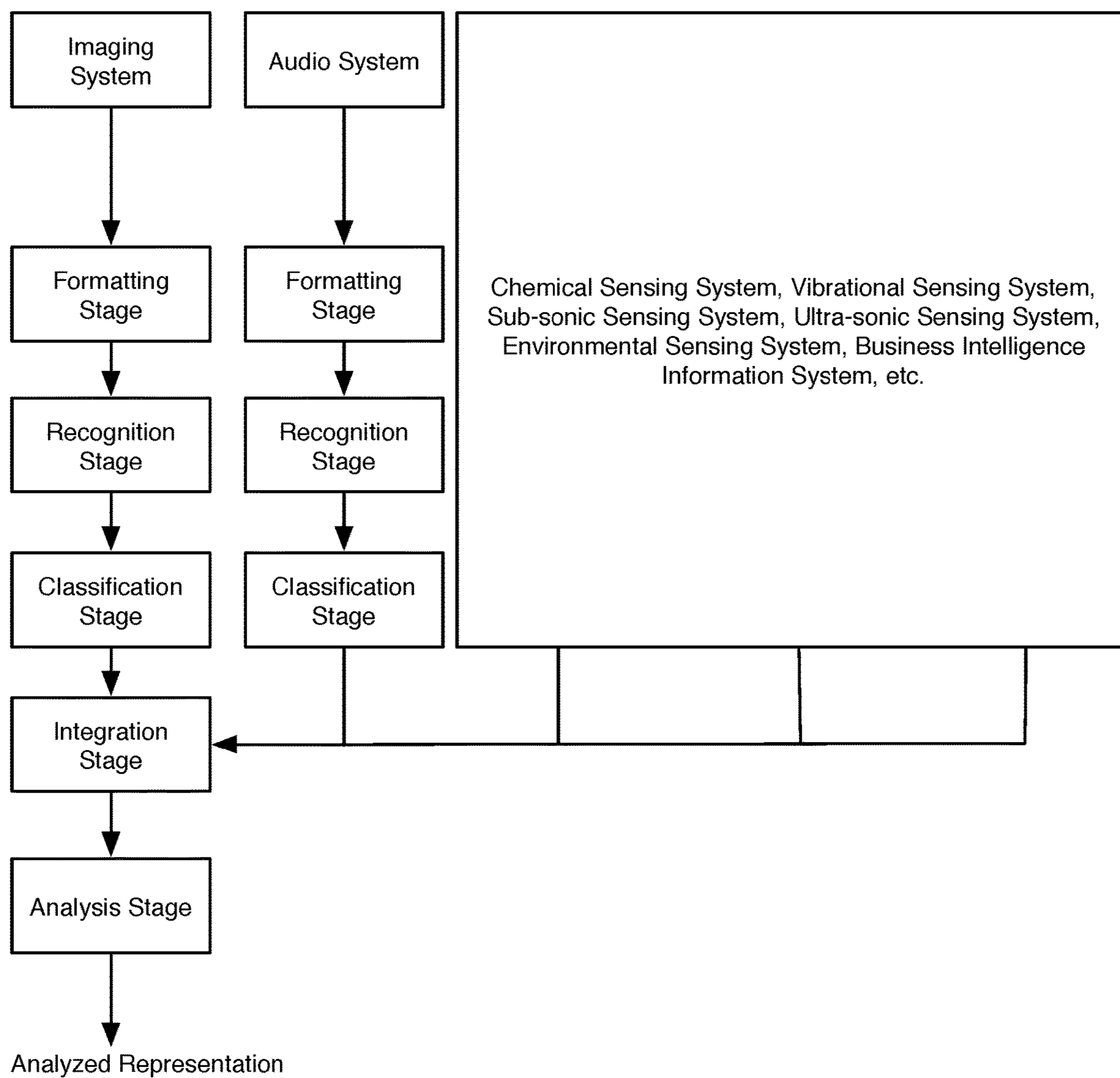


FIGURE 34



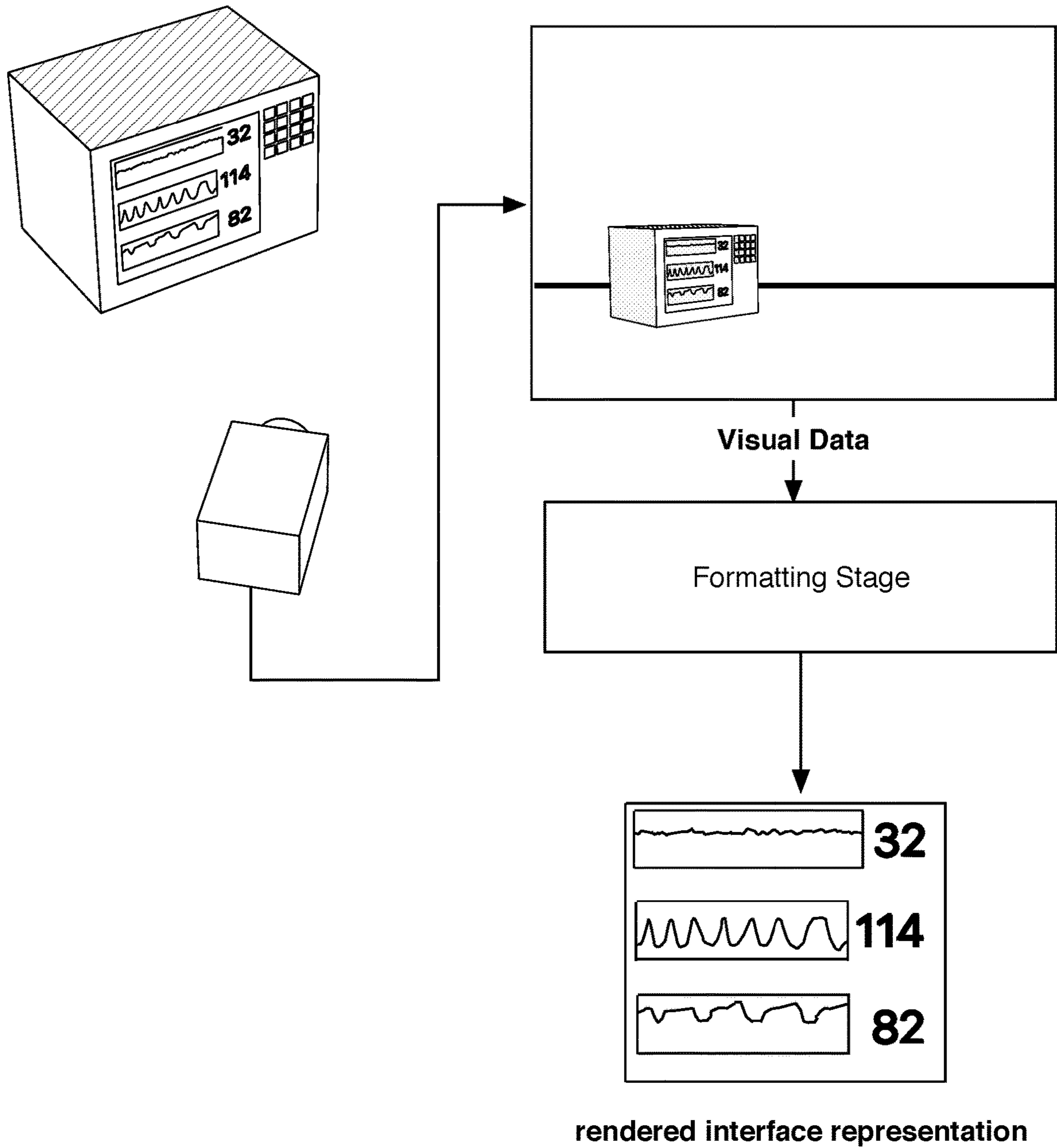
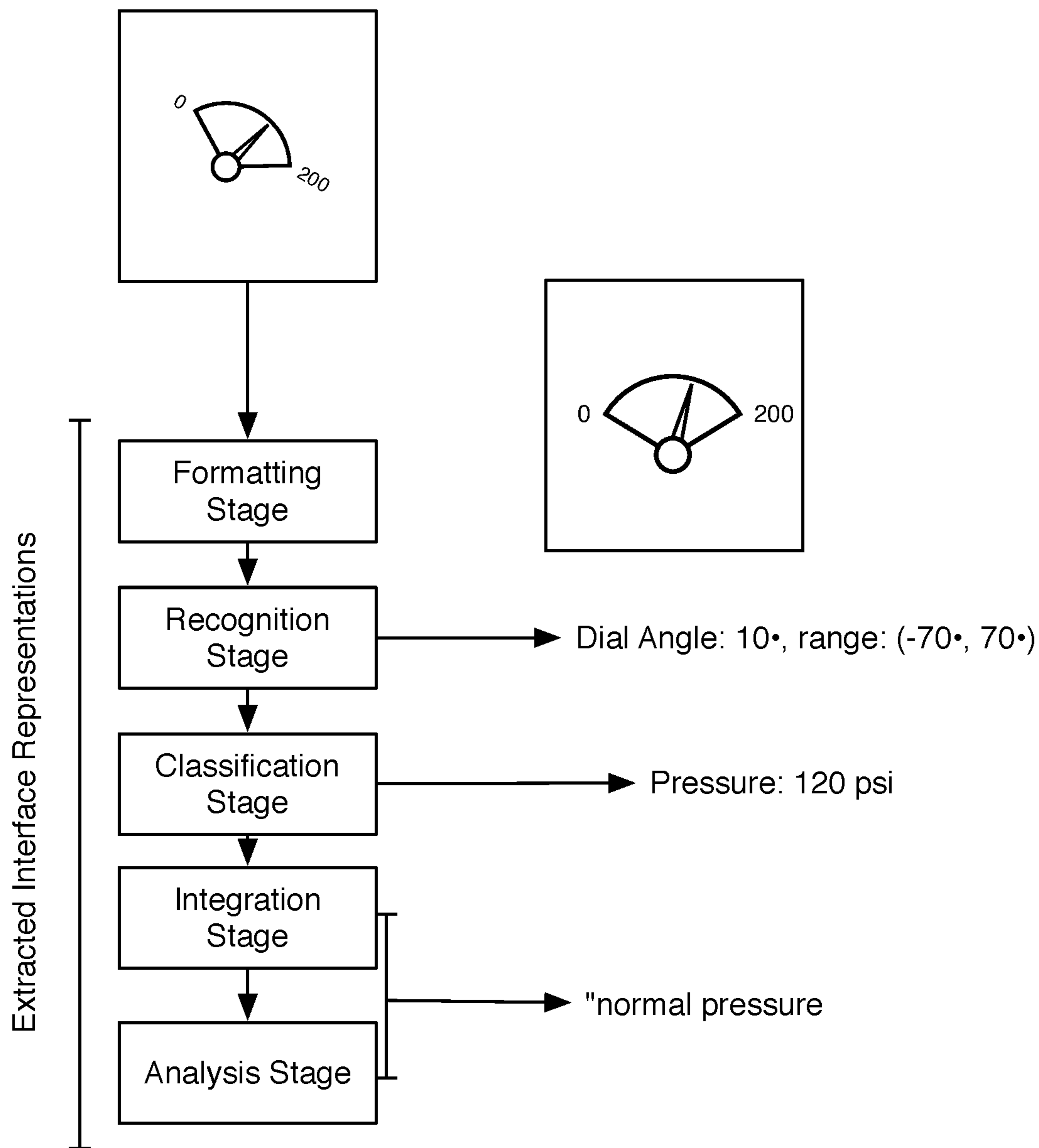


FIGURE 35



**FIGURE 36**

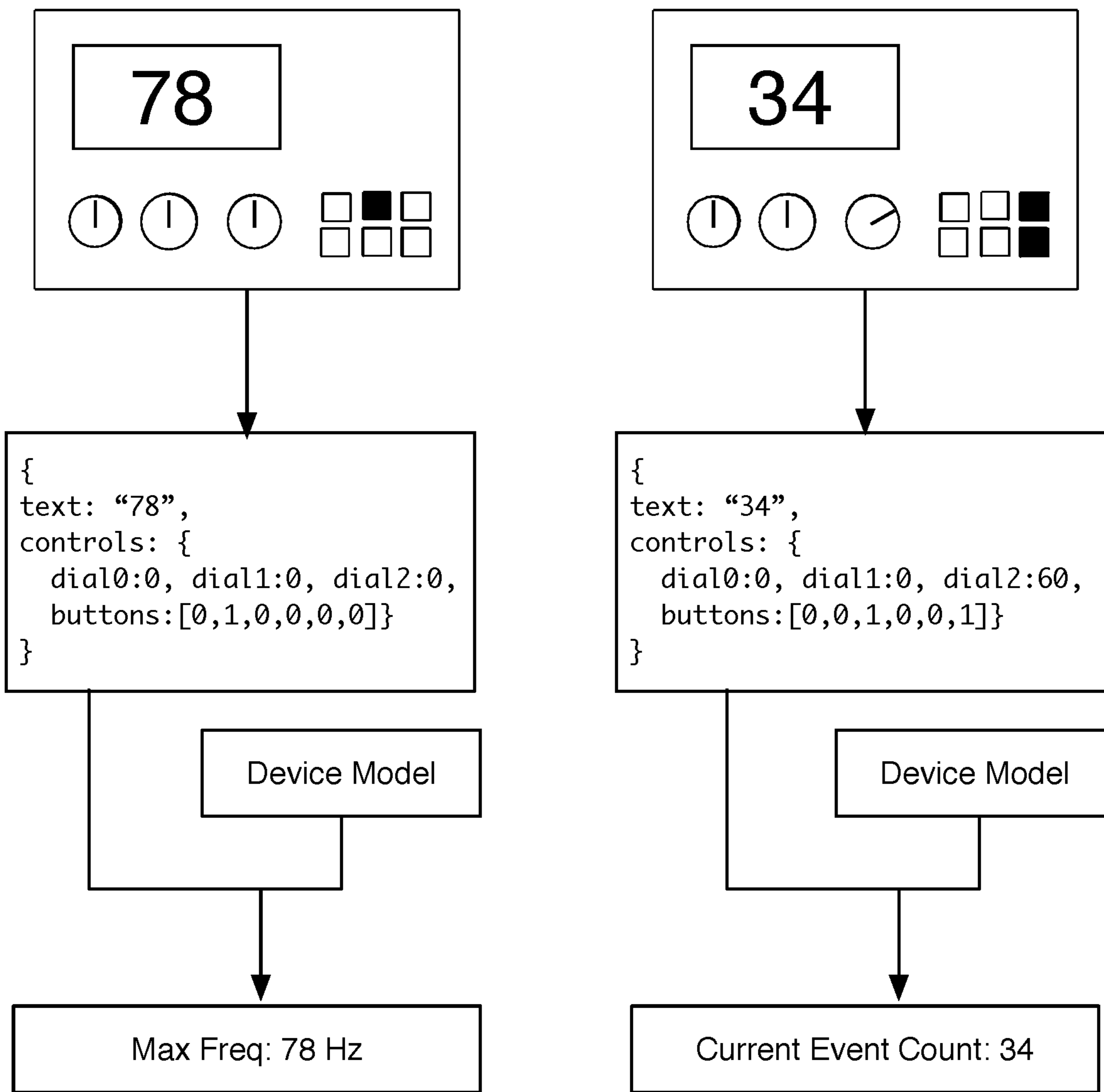


FIGURE 37

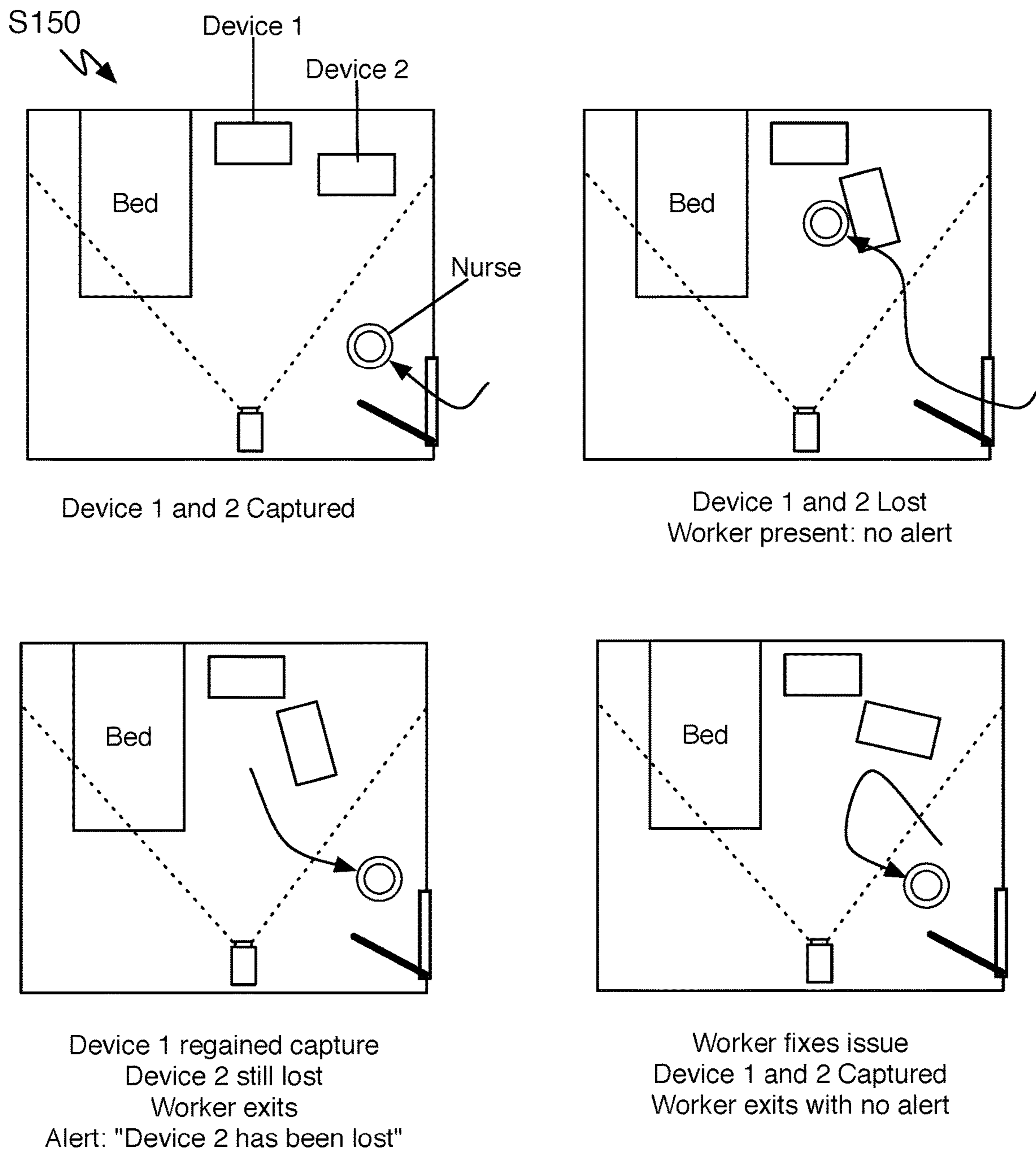
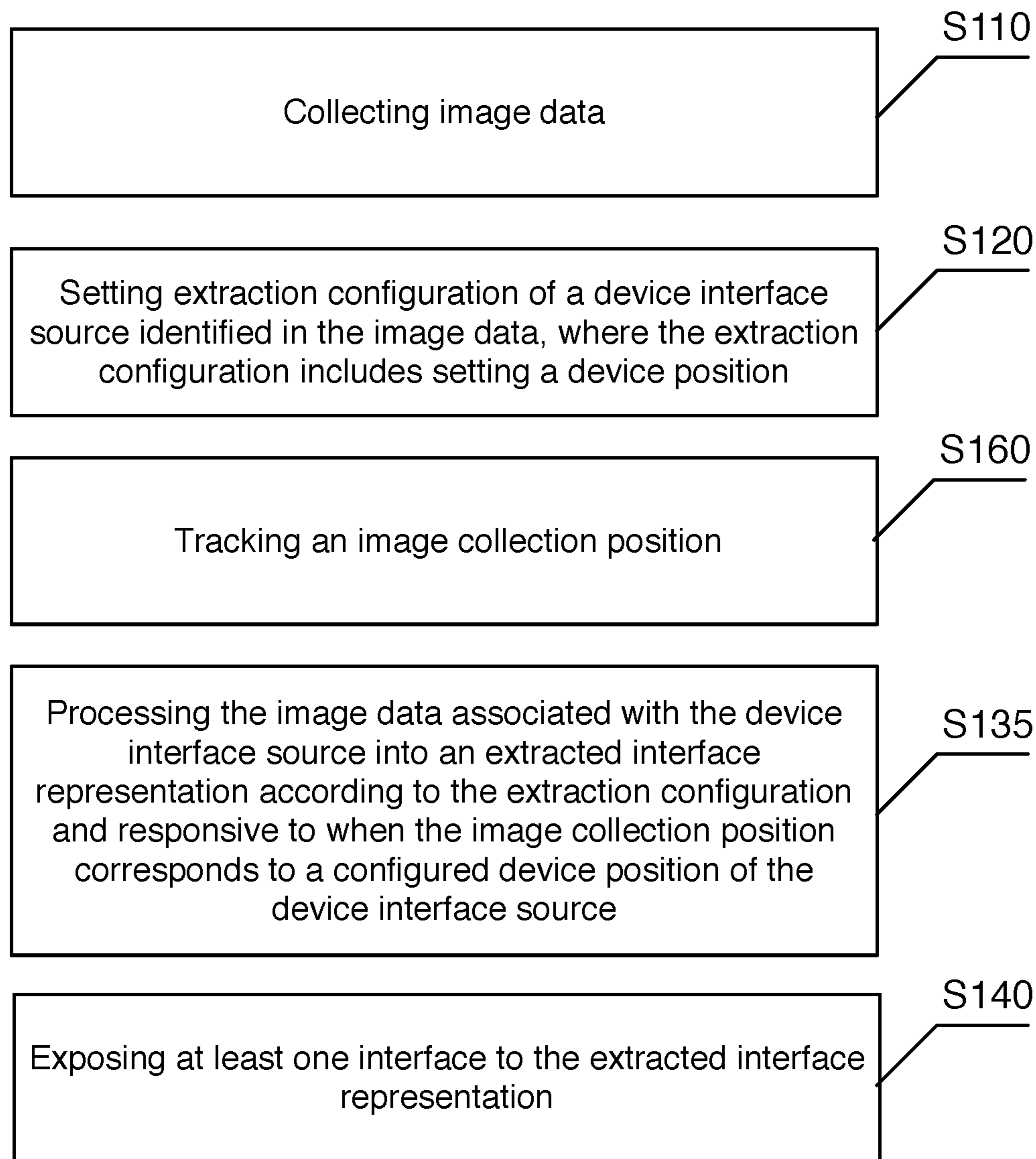


FIGURE 38



**FIGURE 39**



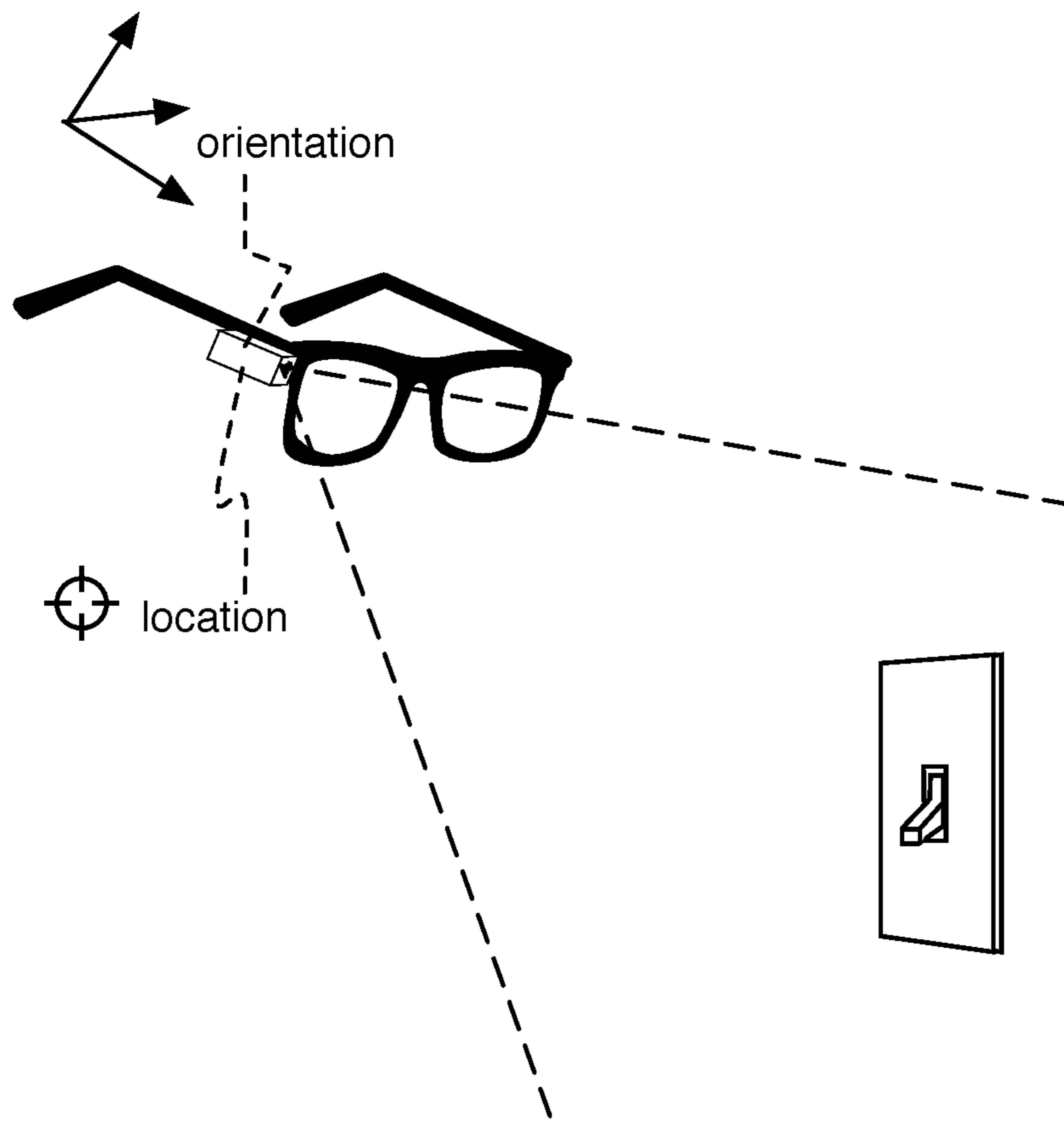


FIGURE 40

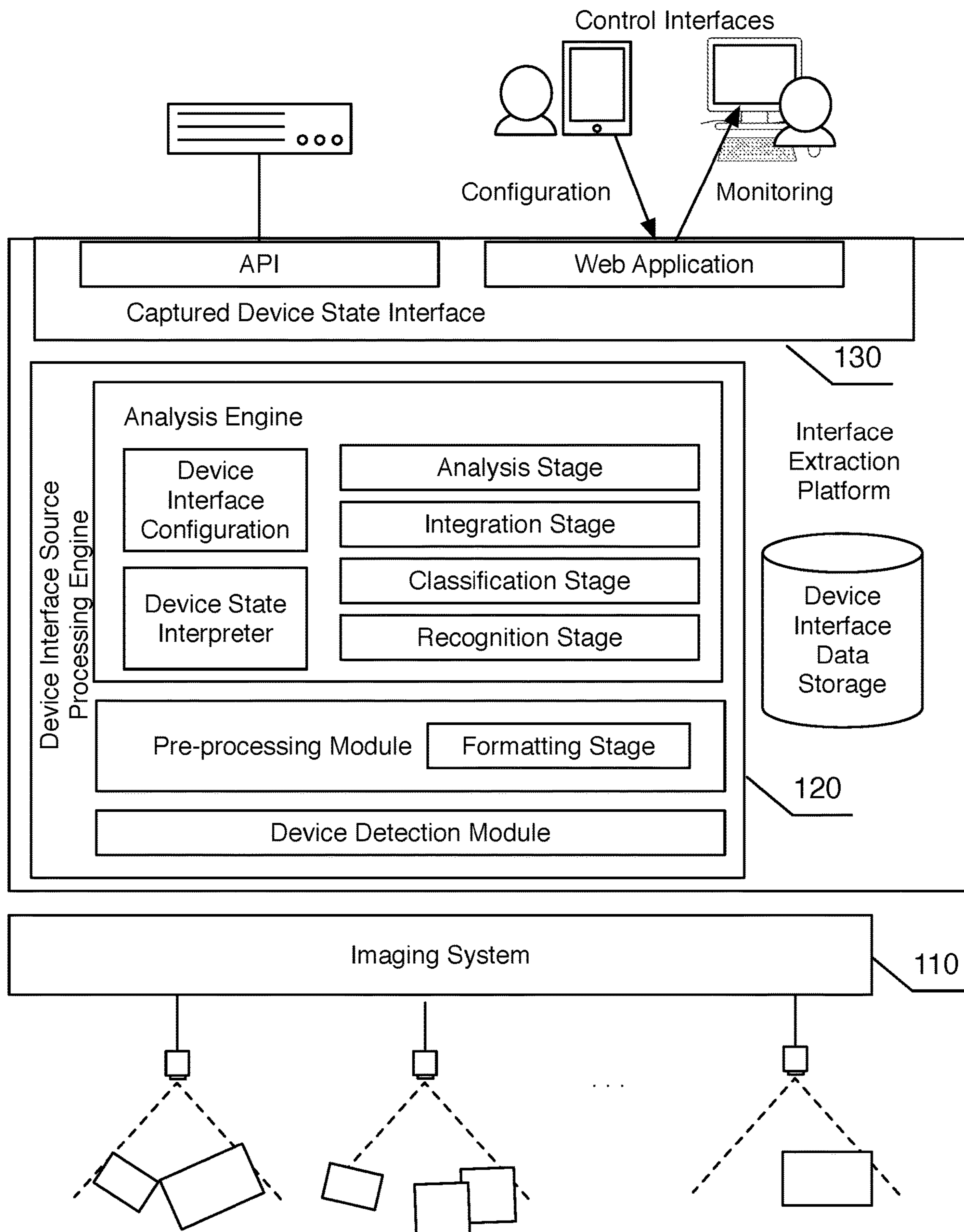


FIGURE 41

**DEVICE INTERFACE EXTRACTION****CROSS-REFERENCE TO RELATED APPLICATIONS**

This application is a Continuation Application of U.S. Non-Provisional application Ser. No. 15/644,799, filed on 9 Jul. 2017, which claims the benefit of U.S. Provisional Application No. 62/360,366, filed on 9 Jul. 2016, and U.S. Provisional Application No. 62/360,369, filed on 9 Jul. 2016, all of which are incorporated in their entireties by this reference.

**TECHNICAL FIELD**

This invention relates generally to the field of device monitoring, and more specifically to a new and useful system and method for remote integration with a device.

**BACKGROUND**

Machines, sensors, computers, and other devices play an integral part in large-scale complex systems. Manufacturing plants, power plants, hospitals, and other large-scale complex systems rely on a variety of machines, sensors, computers and devices as well as humans in forming integrated systems. Many of these systems however lag far behind modern advances and the different devices and systems used within the site operate independently. In some cases, a system uses legacy equipment that lacks any sort of option for data integration. Upgrading equipment can be costly and risky, especially in enterprise situations that depend on the reliability of equipment. Upgrades can also mean downtime time, which is costly or not feasible in many situations. In other cases, a system may have equipment capable of data integration, but the investment to use these device capabilities requires considerable infrastructure investment. This investment is further increased when multiple devices are introduced with their own protocols and capabilities. A fractured Internet of Things (IoT) ecosystem can mean that different companies and products often enforce use of their proprietary cloud hosting environment, use different data protocols, offer different levels of security, have different levels of integration capabilities, provide different data warehousing policies, and/or differ in other ways. It can be prohibitive to build a system for any single connected device and even more complicated building a system to support multiple devices. As a result, many devices, even if they have networking capabilities are left 'disconnected' due to the complication of integrating them. Thus, there is a need in the device monitoring field to create a new and useful system and method for remote integration with a device. This invention provides such a new and useful system and method.

**BRIEF DESCRIPTION OF THE FIGURES**

FIG. 1 is a schematic representation of a system and method applied to a system integration;

FIG. 2 is a schematic representation of an exemplary application of the system and method to interface extraction of a manufacturing device;

FIG. 3 is a schematic representation of the system and method applied to various types of extraction;

FIG. 4 is a flowchart representation of a method of a preferred embodiment;

FIG. 5 is a flowchart representation of a variation of method of a preferred embodiment;

FIG. 6 is a flowchart representation of a method of another preferred embodiment;

5 FIGS. 7 and 8 are schematic representations of different types of imaging device configurations;

FIG. 9 is a schematic representation of a variation where the imaging device is directly coupled to the device interface source;

10 FIG. 10 is a schematic representation of monitoring multiple device interface sources;

FIG. 11 is an exemplary schematic representation of a customization user interface;

15 FIG. 12 is a schematic representation of an exemplary pairing process;

FIG. 13 is a detailed flowchart representation of configuring a device interface source;

FIG. 14 is a graphical representation of the method applied to physical state extraction;

20 FIGS. 15 and 16 are graphical representations of the method applied to indicator detection;

FIG. 17 is a graphical representation of the method applied to switch state extraction;

25 FIGS. 18 and 19 are graphical representations of the method applied to dial extraction;

FIG. 20 is a graphical representation of the method used in a dial extraction mode applied to an analog device interface;

30 FIGS. 21 and 22 are graphical representations of the method applied to character extraction;

FIG. 23 is a graphical representation of the method applied to presence extraction;

FIG. 24 is a graphical representation of the method applied to device model extraction;

35 FIG. 25 is a graphical representation of the method applied to gesture extraction;

FIG. 26 is a schematic representation of an exemplary process using visual markers;

40 FIGS. 27A-27C are schematic representations of visual marker stickers used for configuration and detection;

FIG. 28 is a schematic representation of a device augmentation to convert device interactions into an interface output;

45 FIG. 29 is an exemplary schematic representation of interface output extraction applied to an IV bag;

FIG. 30 is a schematic representation of training extraction of the device interface source;

FIG. 31 is a flowchart representation of a processing pipeline;

50 FIG. 32 is a detailed flowchart representation of processing the image data;

FIG. 33 is an exemplary schematic representation of multiple extracted interface representations;

55 FIG. 34 is a flowchart representation of a processing pipeline utilizing audio and other sensing sources;

FIG. 35 is an exemplary schematic representation of generating a rendered interface representation;

FIG. 36 is a flowchart representation of an exemplary processing pipeline for a dial;

60 FIG. 37 is a schematic representation of interpreting device state according to a device model;

FIG. 38 is an exemplary diagram of generating feedback based on capture status;

65 FIG. 39 is a flowchart representation of a variation of the method used with a movable imaging device;

FIG. 40 is a schematic representation of a movable imaging device tracking image collection position; and



FIG. 41 is a schematic representation of a system of a preferred embodiment.

### DESCRIPTION OF THE EMBODIMENTS

The following description of the embodiments of the invention is not intended to limit the invention to these embodiments but rather to enable a person skilled in the art to make and use this invention.

#### 1. Overview

A system and method for remote integration with a device through a user interface output of a preferred embodiment functions to enable a process for extracting information and data from a device that leverages human accessible and interpretable outputs of the device. The system and method are preferably applicable for applying computer vision (CV) and optionally other forms of perceptual sensing to data acquisition of ambient information in an environment. More preferably, the system and method of preferred embodiments can enable visually automated interface integration, wherein the system and method can enable compatible devices to be automatically detected and interpreted so that information expressed through user-perceptible outputs (e.g., visual and/or audio signals) can be converted to one or more data feeds in the form of an extracted interface representation.

A limited number of modern devices, products, and systems provide data interfaces, but there are a vast number of devices, products, and systems that expose information as usability affordances to human users—most devices designed for use by a human are generally accompanied by interpretable representations of information. However, even in connected devices, much of this information is not exposed for data integration by the actual devices or system. The system and method functions to translate such visually exposed informational manifestations into device data usable in various applications. Accordingly, the system and method can be applied to modern systems as well as legacy systems.

The system and method may be further extended to enable extraction of information from an uncoupled device through any visual and/or audio interface interpretable including machines for visible machine interfaces. In one preferred embodiment, the system and method utilizes a camera to observe a user interface of a device and convert the user interface output into an extracted interface representation. This extracted interface representation can be a formatted data representation of data from the device. Other embodiments can additionally or alternatively utilize a microphone to observe audio-based user interface outputs of a device and convert the user interface output into an accessible interface representation.

In one preferred embodiment, the system and method can be applied to the automated detection of a stateful device. A stateful device can include a device that can present different information at different instances depending on its state. Many medical devices, equipment-operating dashboards, monitoring devices, and the like are stateful devices. A computer, smart phone, tablet, computing kiosk (e.g., a POS system, an ATM, a ticket counter, and the like) and/or other computing devices are also types of stateful devices. The system and method could additionally or alternatively be applied to integration with other basic interfaces such as switches, dials, text displays, indicator signals, and the like.

In an environment where there are multiple areas and devices to be monitored, the system and method can facilitate the creation of a distributed network used for interface

extraction as shown in FIG. 1. In some implementations, the system and method can facilitate achieving full data integration across a large number of devices independent of direct communication with the monitored devices. In a hospital, this may mean that data from medical monitoring devices can be recorded and reported from all patient rooms even as the medical monitoring devices are changed to address individual patient needs or when there is a mixed use of different generations and brands of medical monitoring devices. In a factory, this could mean extracting data from legacy sensors and indicators throughout a complex system without factory downtime. This may be used for analytics within the environment, assisting in surveillance and monitoring, generating alerts and notifications, controlling other systems, assisting with other environment management tasks, or providing any suitable system data integration.

As one potential benefit, the system and method can be used for remotely extracting and exposing the interface output of one or more devices. Remote extraction and visual interface translation enables outside devices to be integrated with the system without altering the outside device or its interface. The device can be unconnected to the system and could be substantially “uncooperative” in the sense that integration with the system is passive from the perspective of the outside device. Additionally, the system and method could be applied to multiple devices and multiple device types. For example, legacy equipment, devices with complicated data interfaces, connected devices with limited data access, and other devices could be integrated with the system. Such extraction modes could be used in combination for a device interface source with analog and stateful presentations of information such as in the exemplary manufacturing equipment shown in FIG. 2. With remote extraction, the device could remain operational during system integration, which could be valuable in industrial situations where equipment downtime is undesired. Furthermore, when applied in industrial settings, well-tested and robust systems can maintain their current designs while still being enhanced for data integration, which may have economical as well as reliability benefits.

As another potential benefit, the system can be used with devices having complex user interfaces. The system and method can manage extraction and exposing of interface output for complex devices with stateful operation. There are many devices offering a wealth of operating and sensing information to humans through graphical user interfaces (e.g., monochrome or color displays). The system and method can account for different display modes and device operating modes in determining what information is displayed and how to extract appropriate information. For example, a medical monitor device may have a small display used to present multiple pieces of information and the pieces of information may be customized depending on how the device is configured. The system and method can account for such customization and appropriately determine and categorize the displayed information as one or more data feed.

As another potential benefit, the system and method can be used for devices with digital user interfaces and/or analog user interfaces. The system and method could interpret and convert user interface elements such as a segmented display, a graphical display, an indicator light, or an analog dial into a digital signal. In some cases, a device interface source may not even be traditionally considered a user interface output but may still convey user interpreted information. For example, the angle of a lock switch can convey the binary state of the lock or the height of a liquid or material in a jar



5

can be a visual representation of content quantity. Accordingly, the system and method could be used with a wide variety of device interface sources including but not limited to: a mercury thermometer, a bi-metal thermometer, an LCD thermometer, an amp meter, a watt meter, a tilt sensor, a shock sensor, a pressure sensor, a flow-rate sensor, a scrolling LED display, a light switch, a circuit breaker switch, a door lock latch, an oven dial, the stick shift of a car, a traffic light, an indicator light, rotary dial, a container of fluid, a container of particulate solids, a handwritten message board, a staging area for objects (e.g., outgoing orders at a restaurant), and/or other sources human interpretable outputs information. Additionally, device designs can be customized for visual interpretation by machines in addition to or in place of humans.

As a related potential benefit, the system and method could additionally be applied for customized creation of data integration with analog systems. In everyday life, people create systems that have informational significance by how they are perceived by people. As examples, restaurants have systems for conveying what to cook and when orders are ready for customers; offices have systems for filing documents and the stage in the processing of those documents; factories and offices use whiteboards to convey information to employees; and families develop systems for reminding each other of chores to name an exemplary set of situations where analog systems convey information. Aspects of these may all be candidates for being used with some implementations of the system and method. Similarly, the introduction of the system and method introduces a paradigm shift in how systems can be designed and the ease with which a digital representation of the analog world can be created. This opens up creative use of the system and method without requiring on customized engineering, algorithm creation, or designs of complex costly systems.

As a related potential benefit, the system and method may accommodate a wide variety of types of generic input sources so that the system and method can be adapted to a variety of data collection applications. There are numerous scenarios where it could be beneficial to have the ability to quickly interface with an existing device—the system and method could offer a set of configuration options such that the system and method could be readily adapted for different applications. Scientists, researchers, engineers, and/or hobbyists could utilize the system and method for custom solutions. For example, an individual may quickly setup a solution for monitoring and tracking oven usage by extracting the oven dial position.

As another potential benefit, the system and method could be used for collecting historical analytics of a device. Data collected through the system and method can synchronize data to a data warehousing system where the data could be accessed, processed, or used in any suitable way. Because of the adaptability to be used with a variety of devices, the system and method can be a tool for enabling the data logging capabilities of an IoT device to a wide variety of devices regardless of the devices own networking capabilities.

As another potential benefit, the system and method can provide enhanced visual monitoring from a remote user interface. In a large integrated system such as a hospital or industrial setting, the various device interface sources can preferably be monitored from a centralized interface. The system and method could be used in formatting and presenting the visual and auditory information in a structured format. For example, a hospital could use the system so that at any time, a nurse or doctor could pull up and view the

6

displays of the monitoring devices in a patient's room. In another example, a large industrial plant could centralize the monitoring of the various devices to a control dashboard. In some variations, data monitoring can be combined with visual monitoring where data can be used to surface image data associated with the site where the data was extracted.

As another potential benefit, the system and method can be used for detection of a current computing context for use of one or more devices in an unplanned situation. In other words, the use and state of one or more outside devices may be detected and used through the system and method. This can be used for providing more in-depth awareness of current situations for user-facing products, machine devices, and/or robots. In hospitals, this may be applicable as different medical devices are often moved from room to room and put to use at different times. The system can preferably adapt to the presence of different devices. In the hospital example, data extracted from medical monitoring devices can be associated with a particular patient, and the continuity of data monitoring for a patient could be maintained even when moving a patient between an operating room recovery room, and/or any suitable room.

Detection of a current computing context could also be particularly applicable when applied with a wearable where the wearable may be used alongside a number of other devices. A mobile variation of the system or method may additionally use automatic enrollment and configuration of device interface sources. As an example, a digital assistant running through the wearable device can be exposed to visually presented information and use that for various tasks. Use of the system and method can alleviate dependence on all outside devices having a digital form of integration with the wearable device. In one example, a doctor, nurse, or worker may use a wearable device to assist in different occupational tasks. The wearable may use the system and method to collect the current computing context so that it can be used in providing contextually aware functionality. In a more general consumer example, walking up to a regular ATM device could result in automatic logging of an ATM transaction performed by that user. In another example, viewing a parking meter can automatically start extract the time from the parking meter and the personal assistant could use that extracted data in starting a time. In both cases, the system and method could be applied without the ATM device or parking meter being “smart” or connected in any way.

As a related potential benefit, the system and method can enable the automatic detection, configuration, and extraction of a device interface source. In a large integrated system, the set of possible device interface sources can be preconfigured to facilitate automatic enrollment. Automatic enrollment may in some cases result in little to no setup wherein data extraction can be activated by installing a camera and enabling the system and method. The automatic enrollment can additionally support interface extraction of multiple devices, which can further enhance operation in a dynamic environment where new machines are introduced, and others are removed.

As another potential benefit, the system and method can provide robust monitoring of the device interface sources. Changes in visibility of a device interface source may alter the confidence in device interface extraction and possibly prevent device interface extraction. Extracted data can be conditioned with confidence levels in one variation. In another, active feedback can be used to assist in correcting and maintaining appropriate visibility of a device interface



source. In time-sensitive critical integrations providing robust data conditioning and collection correction can be beneficial.

As another potential benefit, the system and method can leverage the collection of extracted data with the collection of image data. The image data is primarily used in extracting interface data from a device, but the image data could additionally be used in detecting the environmental context, which can be used to augment or enhance the interpretation of the extracted data. For example, a medical monitoring device may sense various biometric signals, but the system and method could additionally collect the state of the patient as part of the environmental context. In an industrial application, operating conditions of a machine may be associated with detected human workers and/or activity of those workers. For example, an alert system may trigger when certain metrics of a machine, however, the alert system could intelligently avoid triggering if those metrics are changed as a result of a worker changing the settings of that machine. In another application, the parameterized data may be analyzed so that surveillance data can be prioritized for review.

The system and method could potentially be beneficial to a variety of different use cases. In many respects, the system and method provides a unique technical paradigm shift in treatment of imaging and interfaces. Visual interface translation resulting from the system and method can be used as a tool in a wide variety of applications ranging from robust systematized integrations to quick custom integrations.

In some sense, the system and method could be used as a passive sensing solution (i.e., one that can avoid altering operation of the device) that can be setup and configured ad-hoc without directly interacting with a device interface source. In other words, the system and method may not depend on cooperation by the device interface source. To this end, the system and method can be usable across a wide variety of devices, and the system and method may be used in normalizing information into a common format and data interface. Even if a device is a connected device, it may still be easier and faster to integrate with the connected device using the system and method. The system and method could additionally be used in adapting legacy equipment to act substantially like a connected device.

In a preferred embodiment, the system and method may be used in implementing large-scale system integrations. In alternative embodiments, the system and method may be used within various types of products, used in logging device state, used as a tool in building customized CV driven solutions, and/or used with other suitable types of applications.

As discussed, in a preferred embodiment, the system and method can be used within complex environments where there may be multiple different devices and multiple instances of the same device.

One preferred use case would be in hospital settings. A variety of different medical monitoring devices are used in hospitals. A hospital will likely have a wide variety of devices, even for one particular type of device. For example, the hospital may have multiple versions of a heart monitoring devices from different manufacturers and even different models from the same manufacturer. The system and method can be used to extract data from a variety of different devices at one time. Furthermore, the system and method could automatically enroll devices for extraction so that when a doctor or nurse begins to use a new device for a patient, an extracted interface representation of that device can be tracked. Additionally, a hospital could benefit by having improved monitoring of the information from these devices.

Often patients are left in a room with several devices monitoring the status of the patient. Traditionally, only the most critical of devices would be connected for remote monitoring. The system and method could enable more information to be remotely accessible.

The system and method could additionally be applied to industrial use cases that have large, complex integrated systems. Manufacturing plants, power plants, refineries, utility treatment plants, industrial agriculture, chemical laboratory, a fabrication facility, and other industrial settings often have a variety of meters and sensor systems displaying information distributed at different points in the environment. Those various devices could be integrated with the system and method unobtrusively, and potentially with zero down time. Even places of commerce like grocery stores are environments where the system and method could be applied. A store could use the system and method for enabling basic sensing devices throughout the store and collecting data remotely through a surveillance system. For example, scales used in a produce department may display weight through a dial and/or digital display. The system and method may be applied in the store to extract the measured weight of fruit using the existing scale interface. Similarly, the system and method may be used in informational command center installations such as in a cockpit, air traffic control, and/or other environments with multiple displays of information.

As another application, the system and method of an alternative embodiment can be implemented as components of a product such that the product can be specifically designed to leverage visual interface translation. For example, a lamp or lighting fixture can include an implementation of the system and method to enable control of the illumination state by an unconnected light switch observed by a camera of the lighting fixture. Other household appliances, computing devices, electronic devices, and the like could similarly use functionality of the system and method.

As a related application, the system and method of an alternative embodiment may be implemented as capabilities of an imaging device or application. A digital camera, personal computing device with a camera, a wearable computer (e.g., smart glasses), a camera application, a surveillance camera, a robot, an automobile, and/or other suitable devices may leverage the system and method to enable device interface extraction for various applications.

As another application, the system and method of an alternative embodiment may also be implemented as an alternative approach to connected smart sensors. In existing approaches, a device used for logging data is generally produced with a communication module so that it can wirelessly communicate over a network. Not only does this add significant cost to each sensing module, but it also means higher power consumption by the device. It also requires updating existing sensors or devices with the connected smart sensor version. Instantiations of the system and method can be quickly implemented to start extracting data with little adjustment to existing infrastructure. For example, an energy meter could be adapted to a smart meter with logged data by simply having the system and method integrate with the energy meter as a device interface source.

The system and method can additionally or alternatively be used in any suitable use case. The various applications of the system and method may be used independently or in any suitable combination.

## 2. Overview of Types of Extraction

The system and method may be used for a variety of different types of extractions including physical state extrac-



tion, indicator detection, dial extraction, character extraction, presence extraction, and/or device model extraction as shown in FIG. 3. These extractions may be used in any suitable combination and could include other types of extraction. Device model extraction in particular is used in generating an extracted interface representation from a stateful computing device. In some variations, the device model extraction may utilize one or more of the other types of extractions in extracting data, interpreting operating state, or other tasks.

Physical state extraction can comprise a variety of detection approaches. Physical state extraction could be based on color, shape, dimensions, area, volume, appearance, presence, position/location, relative position/location, orientation, or other properties. For example, physical state extraction may be adapted to detect the color of a heating element, and thereby infer its temperature using a black-body radiation curve. Indicator detection and dial extraction are varieties of physical state extraction. Physical state could additionally be applied to graphic interpretation. For example, a line graph generated by some device could be interpreted by converting the graph into a data set.

Indicator detection can be used to detect the binary state or n-state classification of an interface element such as an indicator light, a switch, a circuit breaker, or a lever. Indicator detection can be used for device interfaces with explicit states like a switch, but could also be for natural interfaces like detecting state of a specified lab equipment is in a designated location or classified states such as a the on-off state of a particular medical monitoring device.

Dial extraction can be used to characterize the position of a dial. The dial may be used in signaling information such as a meter. For example, gas meters and/or temperature dials may show measurements through a dial. The dial may alternatively be used as a user input control such as a dial used to set the temperature on an oven. Dial extraction can be used for dials aligned along a radial path or a linear path.

Character extraction can be used to read or detect alpha/numeric characters. Character extraction can be used with digital digit displays (e.g., a segmented display), graphical display of text, printed text, or written text. Character extraction can be customized for a resulting data format.

Presence extraction can be used to detect object presence changes. Presence extraction can additionally be position aware so that an object may be only detected when in a particular region. Alternatively, presence extraction could detect the position and/or orientation of an object or person. Presence extraction could additionally include classification of an object.

Device model extraction can be used to interpret the user interface output of a stateful system such as a computing device with multiple modes of operation. The system and method can handle adjusting visual interface translation in coordination with a currently detected state of a device. This can be used to extract image data from used applications on a computing device. This can also be used for extracting image data from custom device interfaces like medical devices, machine control panels, self-service kiosks (e.g., check-in kiosks, ATM kiosks, etc.), control panels, and the like.

Such forms of extraction can preferably be configured for a particular device interface source. Device interface sources can be configured as being statically located in some location or position. Device interface sources could alternatively be configured for detection within certain regions. In other variations, device interface sources can be configured/processed on-demand through CV-based object classification

and/or identification. Device interface sources could additionally be configured for automatic enrollment/configuration and/or manual configuration, as well as using permissions for restricting capabilities for configuring, processing image data for extraction, and/or accessing data resulting from extraction.

### 3. Method

As shown in FIG. 4, a method for establishing, uncoupled, remote integration with a device through a user interface output of a preferred embodiment can include collecting image data S110; detecting a device interface source present in the field of view of the image data S115; processing the image data associated with the device interface source into an extracted interface representation S130; and exposing at least one access interface to the extracted interface representation S140. The method preferably operates with one or more device interface sources pre-configured with set extraction configuration. The extraction configuration functions to direct the process of how and what data is extracted from the interface. The extraction configuration can specify the type of extraction, where to apply a type of extraction, handling of extracted data, and/or other aspects involved in the processing of image data to generate an extracted interface representation.

The method is preferably used for extracting user interface outputs of a device that would be available primarily to a human user for interpretation, and for converting the related information into an extracted interface representation for remote access, processing, storage, and/or integration. The device interface source preferably includes a visual interface but may additionally or alternatively have audio-based interface elements. The user interface of a device can be an explicit user interface like a display but may alternatively be an analog user interface like the fluid level in a container (e.g., an I.V. bag level), printed text, hand-written text, a knob position, or dial position. The method can be implemented to monitor a single region. The method may alternatively be implemented to monitor a plurality of distinct regions. For example, a hospital may have multiple vision systems installed in multiple patient rooms, and the system can facilitate individually monitoring the various device interface sources in the multiple patient rooms. The method is preferably implemented by an interface extraction system as described below, but any suitable system may alternatively implement the method.

In some variations, the method may additionally include setting extraction configuration of a device interface source S120 as shown in FIG. 5. In some situations, interpretation of a device interface source may need customization in the extraction configuration. In other situations, the interface extraction capabilities of the method may be applied to interface extraction from a device interface source that is not an available pre-configured option. In an alternative embodiment, device interface sources may be more directly configured and may in some cases not use automatic enrollment and detection. This alternative embodiment as shown in FIG. 6 may include collecting image data S110, setting extraction configuration of a device interface source identified in the image data S120, processing the image data associated with the device interface source into an extracted interface representation according to the extraction configuration S130, and exposing at least one access interface to the extracted interface representation S140.

Accordingly, some preferred embodiments of the method may additionally function to enable a user to setup an interface extraction system and adapt the interface extraction system for a particular use case where one or more device



interface sources undergo user interface translation. Preferably, the method performs visual interface translation where graphically apparent information concerning the subject (i.e., the device interface source) The method is preferably implemented in connection with an interface extraction system, which may include a camera to monitor a device and a user application to configure and/or interface with the output of the method. The method is preferably operable on a wide variety of types of extractions including digital and analog interfaces.

The method herein is described primarily from the perspective of visual interface translation of a single device interface source. However, the method can additionally be used for extraction of multiple device interface sources from image data from a single camera, coordinating extraction across multiple imaging devices, performing multiple modes of extraction for single device interface sources, and/or other variations.

As used herein, a device interface source is an object observable through image data that has some properties or output representing some information. A device interface source can be an analog device such as an analog thermometer, a digital device such as a digital thermometer, a stateful computing device, or other suitable objects that convey some information through their state. A device interface source can additionally include an object or objects that traditionally are not thought of as a user interface but represent information. For example, the presence of a coat on a coat rack may be indicative of presence of a user and weather conditions and could feasibly be converted as a data feed through the system and method.

Block **S110**, which includes collecting image data, functions to collect video, pictures, or other imagery of a region potentially containing one or more devices. In one variation, the image data may be directly collected from an imaging device (e.g., a camera) controlled by the interface extraction system. In another variation, the imaging data is collected as a data stream or data file. For example, video may be collected from an existing security system and communicated to the interface extraction system.

Collecting image data preferably includes collecting video data. Collecting image data could alternatively include collecting periodic images or collecting any suitable type of image data. The image data can include high resolution video, low resolution video, photographs from distinct points in time, image data from a fixed point of view, image data from an actuating camera, visual spectrum image data, infrared image data, 3D depth sensing image data, parallax, lidar, radar, sonar, passive illumination, active illumination, and/or any suitable type of image data.

There is preferably at least one imaging device collecting image data. In an implementation within a large environment, a network of multiple imaging devices may be used to form extracted interface representations of different device interface sources. The imaging devices may be used for monitoring distinct regions and distinct devices. The imaging devices may alternatively have overlapping fields of view and, in some cases, monitor the same device. The imaging device is preferably a camera either a video or still camera. The imaging device may collect image data only, but may additionally collect audio, spatial image data, infrared image data, and/or other forms of imaging data. The imaging device in one variation includes an optical system with a wide-angle field of view, wherein generally a given device interface source will be in a sub-region that does not fill over fifty percent of the field of view as shown in FIG. 7. In a wide-angle imaging device variation, locating of the

device interface source may be a process performed when setting extraction configuration in block **S120**. The imaging device in a second variation includes an optical system with a targeted field of view, wherein generally a given device interface source will fill a significant portion of the field of view (e.g., greater than twenty-five percent of the field of view) as shown in FIG. 8. In a targeted imaging device variation, a device interface source may be specified by simply being present or centered in the field of view.

The imaging device may be a stationary imaging device, where the imaging device preferably has a substantially stationary field of view. The imaging device in a stationary variation is preferably positioned to observe an area containing or potentially containing a device of interest. For example, in a hospital, a camera may be directed to the area beside a patient's bed where most monitoring devices are placed. For example, a user mounts or positions a camera at a region to be monitored. In some cases, the device interface source and the scene are substantially stationary. In another variation, the scene is stationary. In another variation, the relative position of the imaging device and the device interface source are substantially stationary. For example, the camera and device interface source may be mounted on a moving vehicle, but the relative position remains substantially stationary. In some variations, the imaging device may additionally be actuated so as to move or be reoriented in different directions. For example, a mounted camera that rotates back and forth. In this variation, actuation position and changes in orientation/position can be accounted for in understanding the position of the device interface source.

A stationary imaging device can be a mounted at a removed/displaced observation point. This can include mounting the imaging device to a spatially removed location from one or more device interface sources as shown in FIGS. 7 and 8. For example, a camera mounted on a wall opposite an oven could track the dials of the oven. In another variation, the imaging device may be an imaging device directly coupled to or positioned substantially near a device as shown in FIG. 9. For example, the imaging device may include a fixturing system to mount the imaging device alongside the visual interface to be extracted. A custom optical system may be used optically account for a non-direct viewing angle. An optical system can facilitate correcting for orientation misalignment of the camera imaging plane and a subject plane, which can mitigate distortion and/or improve focus for regions of interest along a plane of a user interface. A Scheimpflug (e.g., tilt shift) optical system can create a wedge shaped depth of field that can be aligned with the subject plane (e.g., the plane of a user interface display). A directly mounted imaging device may be particularly applicable for high reliability applications. For example, a directly mounted imaging device may cover the face of a gas meter or pressure valve. In such an implementation, the system may include an output display such that collected data can be mirrored on a display (as a data representation and/or a image representation), which can be helpful if the imaging device obstructs the view of the actual output. In another example, the screen of a medical monitoring device may have an imaging device mounted along one border with a corrective optical system as described above.

The imaging device may alternatively be a movable imaging device. For example, images and video collected from a mobile computing device like a smart phone can be used as image data. Alternatively, an imaging device integrated into a wearable computing device like smart glasses or a body camera can provide the image data. In a movable



imaging device, the extraction configuration can be mapped to an image collection position.

In particular for the movable imaging devices but similarly applicable for other types of imaging devices, user accounts may be associated with an imaging device. When the method is implemented as a platform for distributed visual interface extraction, permissions may be used for sharing and cooperative collection of image data. In some cases, image data collected by one user may be used for visual interface extraction of data viewable by a second user and not necessarily the user account that provided the image data. In a workplace environment, multiple workers may be equipped with wearable imaging devices such that they can facilitate the collection of image data for monitoring various device interface sources. In this variation, movable imaging devices may be part of an imaging system network. Movable imaging devices may additionally augment the collection of image data by providing different and variable points of view within an environment.

The above imaging device variations may additionally be used in an imaging device network that collects image data from multiple imaging devices. Preferably, collecting image data occurs from a variety of capture points wherein collecting image data includes collecting image data from multiple image capture devices (e.g., cameras) distributed at distinct points in an environment. The set of capture points can include overlapping and/or non-overlapping views of monitored regions in an environment. The set of capture points can additionally establish a high-density imaging system within the environment. The image data preferably substantially covers a continuous region. However, the collected image data may alternatively constitute non-continuous regions with holes, gaps, uninspected regions, and/or noncontiguous regions. The imaging device network may be used for monitoring distinct regions and distinct devices. The imaging devices may alternatively have overlapping fields of view and, in some cases, monitor the same device. Redundant observation of a device interface can be used to provide enhanced reliability in visual interface extraction, where one imaging device can provide visual observation of a device interface source when another imaging device is blocked. An imaging device network may additionally be used in tracking a device interface source through an environment while periodically or continuously performing visual interface extraction.

The method may additionally be modified to work with auditory interface elements wherein collecting image data may include collecting audio-visual data. Alternatively, collection of audio may be performed separately from the imaging devices. A device may generate distinct sounds to signal different information. These sounds may be characterized by different tone patterns. Auditory data could be collected through a microphone. The auditory sounds can be processed to convert the auditory sounds into an extracted interface representation. For example, a device may emit a particular alarm sound when a sensed metric satisfies some condition, that particular alarm sound could be detected, and a data object could be updated with a property indicating the detected condition.

Block S115, which includes detecting the presence of at least one device interface in the field of view of the image data functions to determine if and where a user interface is observable for extraction. The device interface source is a device or object that has some human-interpretable user interface. Examples of device interface sources can include monitor devices with graphical user interfaces, a dial, a switch, text, a segmented character display, an indicator

light, an analog meter, a printed chart, and/or any suitable visual display of dynamic or static information. The device interface source could alternatively be a general-purpose computer display running one or more applications with a graphical user interface. The image data will preferably include at least one device interface source when the method is active. As shown in FIG. 10, multiple device interface sources can be captured by a single imaging device. In the case where no device interface source is identified, the interface extraction system can periodically check for a device interface source.

The device interface sources are generally distinct from the interface extraction system wherein the vision system is not part of the monitored device. The device may be unconnected (to the system) and may be viewed as “uncooperative” in that it does not alter its operation to accommodate the interface extraction system. Some variations, may have the imaging system and/or interface extraction system mounted directly to the device interface source.

The device interface source could be a device with multiple user interfaces. There could be multiple graphical displays. In some cases, buttons, knobs, and other analog user input elements can convey information. For example, the position of a switch may indicate the operating mode of the device, which may alter the meaning of displayed information. A device interface source preferably has at least one data signal represented in some interface, which can be interpreted as one data signal in a device interface representation. Some device interface sources may alternatively have multiple interpretable data signals that can be included in one or more device interface representation.

Detecting a device interface source can include associating an object with an extraction configuration. The extraction configuration preferably specifies how interface extraction of image data should be processed. For example, detecting of an object with some dial classification may result in selection of an appropriate dial extraction configuration. In the case of a device model extraction, a detectable object can be associated with one of a set of devices models, wherein each device model preferably has extraction configuration pre-set. The set of device models can be customized for different fields of use. For example, in the hospital setting, a set of medical devices used in the hospital can be pre-set with extraction configuration. CV-based object classifier models can preferably be trained to recognize and detect the devices in the set of device models. In one variation, there may be a first subset of device models that are automatically detectable and a second subset of device models that have pre-set extraction configuration but depend on manual selection of the device model. Devices can be visually classified and associated with device models in various ways. General appearance and CV-based object classification and detection may be used. Alternatively or additionally, computer vision detection of device elements like branding and device markings may be used. In another variation, visual markers may be applied to device interface sources of interest, and, when detected in the image data, the visual markers can be used to specify the type of device interface source and the associated extraction configuration. In other variations, a user may manually set the observed device, either before or after detection.

Some variations may include dynamically detecting a device interface source, which functions to automatically enroll a device interface source when it is detected in the field of view. Dynamic detection can be used when multiple devices may be moved into or out of the field of view. In the hospital room example, different patients may require dif-



ferent monitoring devices. These monitoring devices may be wheeled into the room, used for a period of time, and then wheeled out. Dynamic detection could automatically detect the presence of a new device interface source, determine the type of device interface source, and begin generating an extracted interface representation of the device through block S130. Alternative approaches may require an initialization process to setup a device interface source.

The method may additionally support detecting unsupported device interface sources. For example, a generic computing device object classifier could detect general computing devices in the case that the device model cannot be identified. The method could support generating an unsupported device notification. Administrators of the interface extraction system could then facilitate setting of extraction configuration for an unsupported device if desired. In one example, a monitoring interface of a hospital may be notified of an unsupported device interface source, and the administrator could select an appropriate extraction configuration. In some cases, an unsupported device interface source may be detected because the device is unknown, but an unsupported device may also be a detectable device that is in a condition that somehow prevents detection and/or extraction. For example, if something attached to the device prevents proper extraction. In an environment where new equipment is periodically introduced, this can facilitate the maintenance of the interface extraction system by automatically managing monitoring the state of data extraction. There could be cases, where the unsupported device (or a newly detected type of device) may be configured to not be monitored.

As described above, variations of the preferred embodiment may additionally or alternatively include setting extraction configuration of a device interface source identified in the image data S120, which functions to setup a device, interface output, or other contextually-loaded region or object for monitoring. The extraction configuration preferably specifies regions of interest in the image data and how the image data is to be converted into a parameterized data format representing information conveyed through the device interface source. With detection of a device interface source in block S115, the region of interest may be automatically selected based on the detected location of the device interface source.

In one variation, parts of the configuration process may be performed or augmented through user input. In one implementation, a user can use a user interface to configure extraction of the device interface source as shown in FIG. 11. In an instance involving manual setting of extraction configuration, the method can include receiving manual configuration input through a coupled management application. The user interface may display an image from the imaging device and “add” a device for extraction by specifying a position of a device and providing information on how the interface should be interpreted. In yet another variation, augmentation to the device interface source can enhance monitoring of a device interface source.

The setting of extraction configuration may be performed alongside image data collection. For example, actual image data can be used in locating the region of the device interface source. The extraction configuration may alternatively be set independent of image data collection. In one variation, a user interface or configuration file/script may be used in setting the properties of an extraction configuration independent of or without real-time use of image data.

In yet another variation, the configuration may be set through a guided process during a pairing mode of the

system. In a guided process, method may include initializing a pairing mode, identifying a device indicator in the image data during the pairing mode, and associating a device position property of the extraction configuration to a region indicated by the device indicator as shown in FIG. 12. The interface extraction system can preferably be set or requested to enter the pairing mode by a user, and during the pairing mode the system preferably operates to facilitate the setting of extraction configuration. Two preferred variations of identifying of a device indicator can include detecting a user gesture and detecting a visual marker (e.g., a QR code). Other forms of marking a region could alternatively be used.

In the variation of detecting a user gesture as a device indicator, a user may perform some detectable gesture that is detectable through computer vision. Gestures could be static gestures (e.g., thumbs up) or action gestures (e.g., swiping to the left). The user preferably performs that gesture in proximity or against the device interface source. For example, a user may be direct to place their hand, palm open in front of the device interface source of interest and count down using their fingers. The interface extraction system can produce signaling beeps for each count and then confirmation beep when that region is configured for interface extraction. Any suitable framework for gestures may be used.

In the variation of detecting a visual marker, a visual marker may be used to indicate a candidate device interface source. The visual marker could be statically fixed to the device interface source. Alternatively, the visual marker could be a card or some object that a user temporarily positions next to the device interface source. As described below, visual markers can additionally or alternatively be used in various approaches to configuration.

In one variation the initializing of a pairing mode may additionally include setting the pairing mode for a particular type of extraction, which in some variations may simplify the configuration process to identifying the region/location of the device interface source. For example, a user could turn on a dial pairing mode on an interface extraction system, then perform some action to mark the device interface source of interest, and then that device interface source can be configured for extraction as a dial.

In some instances, the interface extraction system may come pre-configured for extraction of one or more device interface sources, in which case the method may not necessitate explicit setting of extraction configuration. As one exemplary use-case, a product using an interface extraction system may be pre-configured during production to recognize and respond to one or more user interfaces. For example, a lamp may come with an imaging device to alter illumination state in coordination with a paired light switch visible in the field of view.

As shown in FIG. 13, configuring a device interface source can include configuring a position of a device interface source S121, selecting at least one of a set of extraction modes S122, and/or configuring device interface operating properties S123. Configuring a device interface source may additionally include registering visual markers on a device interface source S124, and, in some variations, configuring a device interface source may additionally include training extraction of the device interface source S125. Depending on the application, any suitable combination of the processes S121, S122, S123, S124, and/or S125 may be used.

Block S121, which includes configuring a position of a device interface source, functions to indicate one or more locations in the image data associated with the device interface source. Configuring position can set a location and



region of image data. Additionally, configuring position can include setting orientation of the device interface source.

Two potential approaches to configuring a position of a device interface source **S121** can include detecting a device interface source in the image data and/or receiving selection of a position.

Detecting a device interface source in the image data functions to use computer vision to recognize a type of device interface source. Detection of a device interface source will preferably additionally identify a position and region in the image data. In some variations, detecting a device interface may additionally be used in accessing preconfigured default extraction configuration for that type of device interface source. For example, detecting a light switch can automatically result in having basic extraction configuration enabled for that light switch. In a related variation, visual marker may be automatically detected using computer vision and used in setting position and possibly other extraction configuration properties.

Various techniques may be employed in object detection and classification of a device interface source such as a “bag of features” approach, convolutional neural networks (CNN), statistical machine learning, or other suitable approaches. Neural networks or CNNs such as Fast regional-CNN (r-CNN), Faster R-CNN, Mask R-CNN, and/or other neural network variations and implementations can be executed as computer vision driven object classification processes. Image feature extraction and classification is an additional or alternative approach, which may use processes like visual words, constellation of feature classification, and bag-of-words classification processes. These and other classification techniques can include use of scale-invariant feature transform (SIFT), speeded up robust features (SURF), various feature extraction techniques, cascade classifiers, Naive-Bayes, support vector machines, and/or other suitable techniques. Object classification and detection models can be trained on particular types of device interface sources.

Receiving selection of a position functions to have the region set through user input or other forms of input. The selection of the position in one variation can involve user input indicating a position and/or region to associate with a device interface source.

In one variation, selection of the position of the device interface source can be received through a graphical user interface. A user application could display the image data and provide a user interface tool for selection of one or more regions of a device interface source. Such position specification may alternatively be specified programmatically through an application programming interface (API) or other suitable type of interface.

In another variation, selection of the position can be achieved through some interaction during a pairing mode. The pairing mode can be a special mode that when engaged, device interface sources can be identified, located, enrolled, and at least partially configured. Various forms of interactions may be supported for pairing such as a user interaction variation and a visual marker variation.

In a user interaction variation, a pairing mode can include initiating the pairing mode in the interface extraction system (e.g., the imaging device); detecting a user interaction in a region of the image data; and setting the position of the device interface source relative to the region of the user interaction. For example, during a pairing mode, a user could toggle a light switch multiple times to select the region of the light switch as a device interface source. The interaction could additionally be used in training or specifying

other properties. In the light switch example, toggling the light switch can indicate the different states of the light switch.

In a visual marker variation, a pairing mode can include initiating the pairing mode in the interface extraction system (e.g., the imaging device); identifying a visual marker and region of the visual marker in the image data; and setting the position of a device interface source relative to the visual marker. The visual marker may explicitly set the bounds of an inspected region for a device interface source. The visual marker may alternatively be used as user selection of one of a set of possible device interface sources. For example, the visual marker may be placed somewhere in the field of view, and then object detection of viable device using computer vision techniques can be used.

In the variation where the imaging device is movable, configuring position of the device can additionally include setting image collection position, which functions to detect the location and/or orientation of the imaging device when viewing the device interface source. The image collection position may additionally use distance estimation and scene analysis to understand the positioning of the device interface source, which can be used in detecting what image data collected from other perspectives could be candidates for observing the device interface source. The image collection position can include a global location, a local position (e.g., relative to a particular environment), structural position (e.g., room or location in a building or facility), and/or other forms location information. Orientation, direction, and/or information pertaining to the field of view of the imaging device may additionally be included in the image collection position information. A movable camera will preferably have a location service (e.g., GPS, Wi-Fi positioning service, etc.) and/or an inertial measurement unit (e.g., three-axis accelerometer, three-axis gyroscope, and magnetometer) where location and orientation data can be stored as the image collection position.

Selecting at least one of a set of extraction modes **S122** can be used when the method supports a plurality of types of extractions. The type of extraction can be set to customize the visual interface translation for the particular device interface source. The set of extraction modes can include physical state detection mode, an indicator detection mode, a dial extraction mode, a switch extraction mode, a character extraction mode, a presence extraction mode, a device model extraction mode, and/or other suitable modes of extraction.

The different extraction modes may have more granular types of extractions. For example, the physical state detection mode represents one class of extraction modes that can include the indicator detection mode and the dial extraction mode. In the case of a device model extraction mode, a particular device model is preferably specified since a device model extraction mode is generally customized to a particular type of stateful device. The different extraction modes will preferably define different operating parameters in block **S123** and may offer configuration options specific to the particular extraction mode. The different extraction modes will additionally trigger correspondingly different processing routines in block **S130**.

Each position of a device interface source is preferably mapped to at least one type of extraction. In some variations, a single device interface source may have multiple types of extractions configured. In one variation, selection of an extraction mode is specified by received user input indicating one or more types of extraction. In another variation, the method may use computer vision based object detection,



visual marker detection, or other classification approaches to automatically set the extraction mode.

In some cases, multiple types of extraction may be set for an overlapping region of the image data. Some devices may communicate information through a variety of mechanisms, and the method can isolate this information through different types of extractions. For example, a display may display a three-digit number and indicate another property by illuminating the display in one of three different colors—character extraction and color-based physical state extraction can be used simultaneously on that one device.

Block S123, which includes configuring device interface operating properties, functions to set or define parameters that characterize the way a device interface source changes and should be interpreted. The different types of extraction preferably have different sets of operating properties that characterize the components, mechanics, and models for how a particular type of device interface source operates.

Different sets of operating properties may be applicable for physical state extraction, indicator detection, dial extraction, character extraction, presence extraction, device model extraction, and/or any other types or sub-classes of interface extraction. Some varieties may have a small set of operating properties such as setting the value range and metrics of signal dial. Other varieties may have complex operating properties such as device model extraction, which may include multiple state-machines and multiple child device interface source extractions as part of the operating property for one computing device.

Configuring of operating properties can be manually configured, pre-configured to default values, and/or have automatic configuration. Automatic configuration may use computer vision in interpreting and setting one or more operating properties. In the example, of the dial gauge above, the range of values indicated by the dial could be detected through optical character recognition, automatically setting the value range. When being set through at least partial user configuration, the visual rendering of the image data can be augmented with a rendering that visually represents how the device interface source will be monitored. For example, during setting of operating properties, a substantially real-time extracted interface representation can be generated from image data.

Different implementations may expose different levels of operating properties. Advanced properties may expose configuration of low-level interface modeling properties. Higher-level generalized configuration can provide generalized operating properties for common classes of device interface sources. Such generalized operating properties may involve selection of a class of device interface source and then setting of a small set of basic properties. For example, a dial input may be selected for an oven where the temperature range are configurable properties so that based on the angle a set temperature can be extracted. The method may additionally expose a programmatic layer to enable third parties to build out custom logic for a device interface source wherein the operating properties may be defined in a configuration file(s) or application.

As described below, the processing of the image data to generate an extracted interface representation may involve multiple processing stages of formatting, recognition, classification, analysis, and/or integration. Operating properties may be configurable for each of these stages or one of these stages. In some variations, operating properties are not configurable by a user and are pre-configured or automatically set.

As examples of basic operating properties that may apply to many types of extraction, setting operating properties can include aspects such as setting units and setting valid data values, which functions to set the labeling and formatting of the data. Units can be used to label or indicate the meaning of the data. Valid data values can set the space of values expected. This can include setting number value ranges, expected numeral/character output space, value formatting, and/or other properties of valid or likely output from the device interface source. In some cases, the method can support setting of expected operating state to bias the detection to normal operating conditions or conversely make more sensitive to certain conditions. Additionally, the temporal properties of detection and processing such as sampling frequency may be set. Other basic operating properties can include setting different actions or ways of treating. Processing and analysis logic may additionally be configurable such that customized logic can be applied on basic parameterized interface representations. Different types of extractions may use different operating properties for defining the mechanics of a particular device.

Configuring operating properties for a form of physical state extraction can be used to set the sensed properties of one or more objects in a particular region. Physical state extraction can include many various forms of state extraction including detecting general physical state properties, indicator detection, forms of dial extraction, forms of switch or mechanism extraction, presence extraction, and others.

General physical state properties may include setting operating properties that characterize one or more physical attributes of the visual appearance of an object or objects can be monitored in association with a device interface source such as color, size, shape, dimensions, area, volume, appearance, or other properties. As shown in FIG. 14, a user may configure a stove top to be monitored based on the color of the burners. Note that the color range may extend into the infrared in order to better capture color data that can effectively be converted into temperature. Infrared imaging devices can be used for monitoring object temperature, melting of solid, evaporation of a liquid, leaking of a pressurized gas or liquid, paint drying determination, moisture, and/or other aspects.

As shown in FIG. 15, configuring operating properties for indicator detection can include specifying the number of indicated states and classifying at least a subset of the indicated states. The type of indicator could be another property such as illuminated indicator, color indicator, position indicator, shape indicator, and the like. For example, a user may note that an LED indicator light has two states and that the current state is “off”. Image data can be collected and associated with the “off” state. In one variation, image data may be stored and used for visual comparison. Alternatively, a set of image processing routines may be performed to characterize the state algorithmically. The user may note that the other state is “on”. The user may optionally activate the LED indicator light so that image data can be collected on that state. Similarly, an analog indicator such as a light switch or a lock knob could be configured with multiple positional states being configured as shown in FIG. 16. Any number of states can be configured for an indicator.

An indicator can additionally be used as a form of a switch extraction more, which functions to detect the positional state of a mechanism with two or more stable state configurations as shown in FIG. 17. The switch can be a light switch, a lever, or other suitable mechanism. Switch extraction can have similarities to indicator extraction and/or dial extraction. Preferably, the setting the operating proper-



ties includes setting the number of possible states and the respective labels or interpretation of each state. In some cases, some simple state machine logic may be configured as an operating property to specify the possible transitions between states.

As shown in FIG. 18, configuring operating properties for dial extraction can include specifying the minimum range of the dial, a maximum range of the dial, intermediary values, the type of indicator, the path of the indicator, and/or other aspects. The intermediary values can be used to show how values vary along the path of the indicator. Values can vary linearly, logarithmically, exponentially, or with any suitable transfer function. The path of the indicator can be an arc for a radial dial. The path of the indicator may alternatively be linear for a linear dial such as a thermometer. When the dial device is actively used as a user interface output, the indicator is preferably visually distinct. When the dial device is used primarily as a user input, then the indicator may be indicated through the visual appearance of a knob or dial. For example, an oven dial may have a slight protrusion used to indicate its position (as well as serving as a grip for the user). In another variation, the dial may be reversed where the indicator is fixed, and the value rotates as in the oven example shown in FIG. 19.

In some cases, a dial may be a continuous dial that can rotate continuously such as a clock hand. A sampling rate may be set based on the maximum rate at which the dial can rotate so that the method can monitor the continuous dial's progress. The starting position could be initiated when setting up the continuous dial so that subsequent rotation can be tracked.

In some cases, a dial extraction may be configured to indicate progress of an element between at least two positions. Other forms of extraction may enable extraction of position in multiple dimensions. As shown in FIG. 20, the content level in a container could be treated like a dial with the content level acting as the indicator between the top and bottom positions, and the transfer function being proportional to the area cross section at each gradation. This could be used for bottles and/or other containers, where the contents can be treated as a linear indicator along a vertical scale. Deeper operating property customization, such as defining a mapping between indicator position and bottle volume, can enable extraction of volumes.

As shown in FIG. 21, character extraction functions to convert displayed, written, or printed alphanumeric characters into computer readable text data. Configuring character extraction can include setting the region where text is extracted and optionally properties of the text. The properties of the text can include valid characters (e.g., only numbers or only letters A-F), the number of characters, the font or format of the characters, the number of lines, the orientation of the characters, and/or other properties. As shown in FIG. 22, a seven-segment numerical display could be configured to convert the display into numbers representing the current frequency setting of the device.

As shown in FIG. 23, configuring presence extraction can involve setting a background region for detecting presence and setting the conditions for a presence event. Presence extraction can be used to detect when any object enters a space. Presence can be detected by comparing the image data to image data of the background region. The background may be initially set by capturing the image data of the background region with no foreground objects. The method can additionally build an understanding of the background region over time. For example, the method could learn the visual appearance of the background region

under different lighting conditions. Conditions for a presence event can be a set of conditions based on properties of foreground objects. Properties of foreground objects can relate to the size of the object, the shape of the object, the visual texture of the object, the color of the object, or general appearance of the object. In one implementation, image recognition is performed on a foreground object. Presence extraction can be set to indicate if one or more types of objects are present. In one variation, configuring presence extraction can include setting the set of detected object classifications. The set of object classifications may be selected from a set of possible object classifications. In one variation, a customized deep learning neural network model for object classification of a set of objects may be used. Facial or biometric classifications can additionally be used in detecting presence. Presence extraction may alternatively be configured to count the number of objects present currently or over time. Presence extraction can additionally be used with other forms of physical state detection such that color profiles, size, shape, and/or other detectable attributes can be collected in addition to or in response to detection of some object presence.

In some cases, presence detection may be applied in combination with other types of extraction and/or other data collection processes. For example, different tasks or processes may be conditional on detecting (or not detecting) one or more objects. In this way presence may provide context to other forms of data extraction and processes. For example, changes in a light switch may be conditionally extracted when a human is present in the image data.

Presence detection can additionally be applied in a movable device variation, where the location of a device interface source may not be known or static. Presence detection can be configured such that block S130 can enable detecting one or more configured objects and, in response, collecting data (i.e., contextual data). Collecting data in one variation can be performing some type of interface extraction. The interface extraction can be set through extraction configuration that is associated with the object or objects. For example, character extraction can be configured to extract the time displayed on a parking meter when the parking meter object is detected. Collecting data in another variation can include recording of metadata at the time of detecting the at least one type of object and associating the metadata with the extracted interface representation. Metadata may include location, position, position of the object, time of day, audio data, application state of a device, and/or other forms of data collectable from the imaging device, an application/device in connection with the interface extraction system, a third-party service, and/or any suitable source. In one example, presence extraction can be used to detect car keys and to record geographic location upon detection of the car keys. In this way, a user could configure a movable camera (e.g., a wearable camera) to track the last location where the car keys were seen.

As shown in FIG. 24, configuring operating properties of a device model extraction may include setting device model options. In one variation, a number of device models are provided for interpreting stateful computing devices and their respective device states. In this option, configuring the operating properties of a device model may be simply selecting one of the device models and selecting data extraction information options that are of interest for a particular application. Alternatively, a device model can be specified through a device model profile, application, or other programmatic specification of the device states and directions on how to extract information from the image data



during different device states. As a stateful device, a device interface source may present different information in different ways in different conditions. Device model extraction will generally use multiple variations of the above forms of extraction in setting how to extract information data from image data. For example, character recognition in combination with indicator detection may be used to extract different data feeds of information where it is intelligently extracted and associated with a data feed based on detected indicators. Audio interface signals may additionally be used.

As shown in FIG. 25, configuring gesture detection can involve selecting a detectable gesture and/or gesture conditions. In one variation, a gesture extraction mode can be a basic form of presence detection where a gesture is generalized to proximity of two or more detectable objects. Alternatively, particular actions may be trained and detectable a computer vision model. Gestures may be actively performed actions for input, but could alternatively be natural actions that can be interpreted as a gesture. Hand gesture, facial expressions, or other actions can be detectable forms of gestures. Custom gestures, actions, or events could be trained and used for a particular gesture extraction. Gesture detection may be reactive to presence detection. However, gestures may alternatively be linked to detection at a particular position. In one exemplary application, the gesture of locking a door lock with a key may be a detectable gesture. The physical action of locking may be detectable through computer vision. Alternatively, gesture detection could be composed of detecting presence of a person, a key, and the lock within some threshold of near proximity and/or making contact.

Block S124, which includes registering visual markers on a device interface source, functions to use visual markers in facilitating configuration. The visual markers could be stickers, marks, or attachments that can be physically added to the device or otherwise augment how the device is perceived. The visual markers are preferably visually distinct and made of a distinguishable color and/or graphical pattern. Additionally, the visual markers could have a machine readable code such as a QR code or a barcode. The machine readable code is preferably set to an identifier. The identifier can be used to indicate different things such as the purpose and meaning of the marker and how it should relate to extracting information.

Two preferred variations of using visual markers can include using the visual markers in specifying at least a portion of extraction configuration and/or using positioning of visual markers in setting at least a portion of operating properties. As discussed above, a pairing mode is another potential application of visual markers, but they may alternatively be used in other ways.

In a variation where visual markers are used to specify at least a portion of extraction configuration, visual markers may be placed on different device interface sources where the visual markers indicate extraction configuration. Detection of a visual marker can indicate position of a device interface source but may additionally indicate an extraction mode. Accordingly, the method can include providing a visual marker associated with at least one extraction mode, detecting the visual marker in the image data and setting at least part of the extraction configuration based on properties of the visual marker, wherein position, extraction mode, and/or device interface operating properties can be set as shown in FIG. 26. In one variation, the visual markers may visually represent an identifier. The identifier can be used in accessing extraction configuration associated with that identifier. The identifier can be unique such that the extraction

configuration can be uniquely customized. The identifier may alternatively not be unique where it can reference an extraction configuration that can be shared across multiple device interface sources. Alternatively, the visual markers may represent some visually encoded message that embeds the extraction configuration properties. In this variation, the visual marker may be used without needing access to a database of associations. In this variation, common extraction configurations may be associated with visual markers that can then be positioned or attached to different device interface sources of interest. Image data containing that capture the visual marker will then be automatically processed using the extraction configuration.

In another variation, visual markers may be used in marking different aspects to assist in the extraction of information. This variation preferably leverages different visual markers to augment a device interface source so as to make it more easily or reliably interpreted. This variation may include providing a set of visual markers, such that the combined positioning of the set of visual markers as detected in the image data at least partially define extraction configuration of the device interface source. Visual markers may have attributes to specifically configure particular aspects of extraction configuration. Different types of visual markers may be used to configure the position of a device interface source, the type of extraction, and operating properties as shown in FIGS. 27A-C.

A position setting visual marker may be used to mark where a device interface source is located and possibly set the dimensions of associated image data. A position visual marker may be a box that can be placed to circumscribe a device interface source. Other systems of a visual marker could alternatively be used.

An identifier on the visual marker may indicate the type of extraction to be used. Element visual markers may be used to characterize the operating properties of the device interface source.

Operating property indicating element visual markers may be used to specify different operating properties. The positioning of a property visual marker can convey some information used in setting an operating property, the type or label of an operating property visual marker may additionally be used. For example, a min-value visual marker and max-value visual marker may set the range of a dial. An operating property visual marker may additionally include variable fields where information can be included to set variables. These may be used in combination to set multiple operating properties. For example, dial-based visual markers may visually mark the minimum value, intermediary values, maximum value, orientation of an indicator, path of the indicator, and/or any suitable operation property.

Once placed, visual markers are preferably detected and used to automatically configure aspects of the device interface source. The visual markers can act as a physical approach to normalizing detection of different types of devices. They may additionally be used to artificially make particular device(s) be perceived as a different type of device. For example, dial extreme markers can be placed on a cone and wall of a garage with another marker on the front of a car to convert a car pulling into a garage as a progress bar dial as shown in FIG. 28.

In another exemplary application, the method could be applied to an intravenous therapy bag (IV bag). The IV bag could be a common IV bag such as a drip or pump system, which enables existing equipment to be used. The IV bag may be augmented to enhance the fluid level presence. In one variation, a light could be used to illuminate the fluid to



enhance detecting its level. In another variation shown in FIG. 29, a graphical pattern could be placed in the background. The optical disruption of the pattern caused by the fluid could be differentiated from disruption caused by just the bag, and the fluid level could be derived from this observation.

Visual markers could similarly be applied to other applications of monitoring liquid contents of an at least semi-transparent container. For example, bottles of liquid could similarly be monitored. A visual marker strip could be adhered to the back of a bottle. Alternatively, a visual marker background could be positioned such that the bottles of interest are the visual marker background and the imaging device. The visual graphical pattern in this variation may be a graduated pattern with continuous or discrete segments of different identifiable levels. The level of the liquid can correspond to marked liquid levels based on the detection of a graduated pattern.

In a similar variation that does not utilize augmentation the method could learn the background visual representation and use a similar approach to optical disruption to identify the fluid level.

Block S125, which includes training extraction of the device interface source, functions to use machine learning in understanding the informational classification of image data as shown in FIG. 30. Training could be performed during an initialization process, but may additionally or alternatively be performed during operation. Training preferably involves collecting a training set of image data and generating a model for associating the image data of a device interface source with various values. The training may be beneficial in scenarios where the device interface source does not perfectly map to one of the available types of extractions. Training may be an alternative approach to configuring. For example, instead of setting up how a dial works, a user could show the imaging device a number of possible dial positions and input their associated values. The interface extraction system can alert the user when enough samples are collected. The interface extraction system does not need to have samples for a full set of possible values because, the method could interpolate the visual detection of un-trained values. The training and automatic detection of device state can be performed using computer vision, deep learning, neural networks, and/or other suitable forms of machine intelligence.

Block S130, which includes processing the image data associated with the device interface source into an extracted interface representation, functions to organize, characterize, or convert the image data into a format for consumption. The processing of image is preferably performed according to extraction configuration of the detected device interface(s). As discussed, one or more device interface sources can be extracted simultaneously from the image data. Additionally, one device interface source may have one or more types of extraction applied to the image data of that device interface source. In some cases, this may result in multiple data streams of extracted interface representations. An extracted interface representation is preferably a data object that reflects the state of the device. Various embodiments of the method may provide one or more types of representations such as a rendered interface representation, a parameterized interface representation, a classified interface representation, and/or any suitable analyzed representation.

A rendered interface representation can be a virtual rendering or an image-based rendering of the interface. The rendered interface representation can provide a visual representation. An image-based rendering can be a formatted

version of the device interface source as captured in the image data. An image-based rendering may be saved and used to provide the base truth of the state of device interface source. A virtual rendering can be a simulated visual representation that can be constructed from a parameterized or classified interface representation. A rendered visual representation in some applications can be streamed or provided as a visual representation to accompany a machine translation of the information.

A parameterized interface representation can be a machine-readable representation of the interface. In some variations, an initial form of parameterized interface representation can be a conversion of image data of a device interface source to a data representation. This can include reducing the components of the interface into its base components. An example for a dial device interface source would be the dial angle and position relative to the dial range. Those parameterized representations of the base components may then be converted to a classified interface representation. A classified interface representation is preferably achieved through characterizing informational data from a base parameterized representation. In some cases, processing of the image data may bypass any intermediary form by using image classification, feature extraction, or other CV-based approaches. Alternatively, the parameterized interface representation of base components may be used as inputs used in generating the classified interface representation.

The classified interface representation, or in other words an informational data representation, is a type of a parameterized interface representation data format that provides the symbolic meaning of the interface and is generally what an informed user would interpret from viewing the device. In the dial example used above, the classified interface representation would be the indicated value of the dial.

An analyzed representation can be a higher order data information format resulting from applying high level logic, statistical analysis, and/or understanding of the information conveyed in the device interface source. The analyzed representation can be the result of applying high level logic and processing of the information as applied to the extracted interface presentation, and may use historical data, other data sources, and/or other resources in performing the analysis.

One or more of these extracted interface representations may be exposed as shown in FIG. 31. In one preferred implementation, the extracted interface representation can expose a classified interface representation and the image-based rendered interface representation so that the ground truth of the classified interface representation can be inspected.

The extracted interface representation is preferably stored in association with a timestamp and can be periodically updated such that a longitudinal history of an extracted interface representation can be established. In some cases, however, a single or set of isolated snapshots of an extracted interface representation may be produced. As the extracted interface representation may not be collected with at regular intervals, data flagging or eventing may be triggered in response to issues in generating an extracted interface representation.

The processing of the image data may be scheduled such that it is executed at set intervals. The processing of the image data may alternatively be responsive to detection of the device interface source. A device interface source detection process can be executed continuously or periodically on the image data, and upon detection of the device interface



source, the image data can be appropriately processed. This can be particularly applicable to variations using a movable imaging device, where the device interface source may not always be present or located in the same region. The processing of the image data may alternatively be triggered through some input. For example, when integrated with a digital camera, the processing can execute after capture of the image data. In another example, a programmatic command may be communicated to an imaging device to capture image data and process the image data.

As shown in FIG. 31, processing the image data can include a number of processing stages, which may involve a visual formatting stage, recognition stage, classification and data formatting stage, and/or analysis stage. There may additionally be one or more integration stages where outside data can be used to augment any one of the stages. Accordingly, processing of the image data can include extracting image data associated with the device interface source S131, formatting the image data S132, extracting the interface output of the device interface source S133, and/or analyzing the extracted interface output S134 as shown in FIG. 32. Various embodiments may only make use of a subset of the possible stages. The various processing stages could additionally be used to generate different interface representations including a rendered interface representation, a parameterized representation, a classified interface representation, and/or an analyzed representation as shown in FIG. 34. Each stage of processing may output a different extracted interface representation. Additionally, the processing pipeline may utilize information and processed data from other sources and/or sensors as shown in FIG. 35. In many system integrations, the various representations may each be exposed and used in different ways. For example, the rendered interface representation can be used for remote visual inspection, the classified interface representation can be used for historical logs, and an integration/analysis stage can be used for controlling some other system.

Block S131, which includes extracting image data associated with the device interface source, functions to isolate the image data associated with the device interface source. Multiple cropped segments of image data can be created from the image data if multiple device interface sources are present.

In some variations the position of the device interface source is substantially static and configured in the extraction configuration. The selection of a sub-region of relevant image data can be directly performed based on the configured position of the device interface source.

In other variations, block S131 may include detecting a device interface source and the position of the device interface source. This may be achieved through object recognition using CV-based approach. This could alternatively be achieved through detection of a visual marker or other suitable approaches.

Block S132, which includes formatting the image data, functions to prepare, normalize, and/or rectify the image data as part of the visual formatting stage. Preferably, formatting of the image data includes applying image data transformations to a region of image data and thereby generating formatted image data of the device interface source. In some variations, the formatted image data prepares the image data to be used in subsequent stages of processing or for a rendered interface representation. Formatting the image data may additionally be used in generating a rendered interface representation that can be part of the extracted interface representation output. In some variations, the image data transformations of image data relevant

to a device interface source may substantially make up the final extracted interface representation outputted by the interface extraction system. In one implementation, video and/or images of various device interface sources may be formatted into a rendered interface representation and streamed for easier browsing and remote viewing as shown in FIG. 35.

Various visual transformations can be made including geometric transformations and imaging transformations. In one basic form of formatting the image data, the image data can be cropped to a region of interest for the device interface source. For example, the image data may be converted to one or more sub-images that are cropped to some bounding box around the device interface source or some region of interest for the device interface source. The visual transformations are preferably specifically customized for regions of image data associated with different device interface sources such that each device interface sources can be normalized into an extracted interface representation. If there are two or more device interfaces sources captured by a camera, each device interface source can be transformed individually. Image data transformations are preferably applied directly on the image data, but can additionally include altering the configuration of an imaging device to transform image data through altering the image data collection process.

Geometric transformations can rotate, skew, distort or otherwise morph the geometry of one or more regions of the image data of an interface to account for perspective. Accordingly, the viewing angle of a device interface source can accommodate rotation and non-normal viewing perspectives. For example, screens viewed off angle are transformed to remove key-stoning in order to produce regular rectangular images with aspect ratios that might match that of the screen of the original device. The amount of visual transformation can additionally be used to characterize the quality of the collected data (i.e., the capture status). For example, a screen viewed from an extreme angle (e.g., eighty degrees off of a straight-on view) may be less reliable. Reliability of the source of the extracted interface representation can be stored as a quality data property. The data property can depend on the base image data and/or other aspects such as confidence in translating image data to a parameterized value.

Other visual transformations can include imaging transformations, which functions to adjust the color space of the image data for enhanced legibility and/or processing. As with other image data transformations, imaging transformations can be customized to different regions of the image data. Imaging transformations can include adjusting the color space, brightness, contrast level, saturation level, hue, sharpness, white point, black point, and/or altering any suitable imaging variable. Filters or other image transformations may additionally be performed. The type of imaging transformations can be based on the type of extraction. The type of imaging transformations can be based on the classification or other understanding of the device interface source. For example, if a medical monitor device displays a large digit representation of a vital sign and a graph of a second vital sign. The region associated with the first vital sign may be adjusted to improve readability of the first vital sign, while the region associated with the second vital sign may be differentially adjusted to improve readability of the second vital sign. When a targeted device has a color display this could include such imaging transformations such as adjusting color balance in one sub-region of the display displaying data with one color profile and a using a second



imaging transformation for a second sub-region displaying data in a second color profile.

Block S133, which includes extracting the interface output of the device interface source, functions to convert image data related to the device interface source into some form of a parameterized interface representation. A parameterized interface representation preferably includes a classification and data formatting stage wherein a classified machine readable representation of an interpretation of the device interface source is produced. For example, the image of a dial can be converted to the interpreted meaning of the observable dial setting. In some variations, extracting the interface output may additionally include a recognition stage where components of the device interface source are detected. For example, a parameterized description of the position of the dial and the range of the dial in a generic description.

In a recognition stage of extracting the interface output, visually represented information is preferably converted to a base parameterized interface representation. Accordingly, processing the image data may include extracting a parameterized representation of the formatted image data. The parameterized interface representation can be used in providing a lower-level interpretation of the device interface source in between performing image processing and obtaining a classified interface representation. A base parameterized interface representation is a parameterized data representation of a visual description of the interface. In one variation, the parameterized interface format of a device interface source can be a data object with an array of graphical elements where each graphical element can have detected text, color properties, a location property, size properties, orientation properties and/or other properties. The properties of a data object for a base parameterized interface representation may depend on the type of extraction. For example, a segmented character display can include a first set of properties and a dial can have a second set of properties. Alpha/numerical characters are preferably recognized through optical character recognition (OCR) techniques or other processing techniques. Visual representations such as status indicators, progress bars, graphs, and other interface representations of information can be converted to a normalized parameterized interface representation. A dial may be represented as an angular or linear position of the dial indicator as shown in FIG. 36.

In a classification and data formatting stage of extracting the interface output, the nature of the device interface source is interpreted into a classified interface representation also referred to as informational data. The classified interface representation is a data format that provides the symbolic/informational meaning of the interface and is generally what an informed user would interpret from viewing the device. A classified interface representation is preferably achieved through characterizing informational data from a base parameterized representation. In some instances, or variations, a base parameterized representation may be used to determine the symbolic meaning of the interface output. Alternatively, the image data may be used directly. In instances where a device interface source has two or more data signals, characterizing a classified interface representation can include characterizing two or more sets of classified interface representations. For example, a medical monitor device may present two data signals as a real-time value represented as an alphanumeric display and a third data signal as a line graph.

Depending on the extraction mode, a classification and data formatting stage may include various processes such as

interpreting alphanumerical representation of an interface output and/or interpreting graphical representation of an interface output.

In the rotating dial example used above, the angular position of the dial may be converted to the indicated value after determining that the dial represents zero at negative sixty degrees and two hundred at sixty degrees as shown in FIG. 36.

In an alternative approach, an algorithmic approach applying machine learning, neural networks, image classification, and/or other techniques can be used to map the visual data associated with a device interface source to a symbolic representation. For example, the image of a dial may be classified as having a higher correlation to a set of images known to have a particular symbolic mapping.

A base parameterized interface format may be used to determine the symbolic meaning of the interface output. In one variation, a parameterized representation of the image data can then be characterized into a classified data representation, which provides more informational data that reflects the conveyed information as opposed to a parametric description of the appearance of the interface. Alternatively, the image data may be used directly such as when using computer vision and machine intelligence to classify the meaning of the image data. The base parameterized interface format preferably represents the main components that define the mechanics of how the device interface source operates. The base parameterized interface representation can be logically analyzed to generate the classified interface representation. For example, a dial may have a minimum value representing zero at a first location, a maximum value representing 100 at a second location, and an indicator currently at a third location halfway between the minimum and maximum. This exemplary parameterized interface representation can be converted to a classified interface representation of fifty. The base parameterized interface representation may alternatively be used as input for machine learning analysis.

The units assigned in the classification stage may be dynamically determined from the presented interface output but could alternatively be pre-configured during block S120. For example, a user may specify that the number characters read on a digital character display represent the temperature in Fahrenheit. Additionally formatting rules can be set to appropriately format the parameterized interface representation.

In one implementation for a type of extraction, computer vision classification may be applied in translating from image data to an extracted interface representation and more specifically a classified metric representation. As one potential solution for extracting an interface representation, the method may apply deep learning, neural nets, and/or other forms of machine learning models that are trained to translate image data of a type of user interface to its interpreted meaning. The interpreted meaning can be normalized and then extraction configuration can be used in scaling that to the intended interface output. For example, the positions of a radial and/or linear dial may be trained so that a normalized position or angle of the dial can be detected through a neural network. Then the configured scale of the dial (e.g., linear/logarithmic, value range, etc.) can be used to convert the normalized position or angle to a data value. Other interfaces may be interpreted through deep learning where the visual appearance has a mapping to interpreted meaning. In the example above, the image of a dial may be classified as having a higher correlation to a set of images known to have a particular symbolic mapping.



The various types of configurable extraction modes may have specific processing steps customized to interpreting each type of interface.

In a physical state detection mode, processing the image data associated with the device interface source into an extracted interface representation can include parameterizing visual physical state of the device interface source into an extracted interface representation. Parameterizing the visual physical state can include calculating size, tracking shape, tracking color profile, tracking orientation, tracking position within a region, and/or tracking other attributes of the physical state of an object. In some instances, the device interface source is static, in which case the physical state can be processed when the device interface source is visible. In other instances, the device interface source may apply physical state detection upon detecting presence of a particular object. The physical state properties that are parameterized may be conditional based on the present object in a region of the device interface source.

In an indicator detection mode, processing the image data associated with the device interface source into an extracted interface representation can include parameterizing the indicator state of at least one indicator signal. One preferred type of indicator is lighted indicator signal in which case the illumination state is parameterized, which may include parameterizing stable state of an indicator, parameterizing time varied states, parameterizing illumination qualities, and the like. This can include detecting an on and off state. There may additionally be other states. In one variation, the indicator state may be communicated through state variations over some window of time. For example, a certain number of flashes or rate of flashing can be detected and mapped to a configured state. Color, brightness, and other illumination qualities of an indicator may also be detected and classified as different states. An indicator detection mode may additionally be used for physical indicators like switches, buttons, graphical indicators. In one variation, multiple indicators may as a collection indicate different information, and parameterizing can include parameterizing the collective state of the indicators. For example, three indicator lights may turn on and off in different combinations to signal different information which could be interpreted through logic in the extraction configuration.

A switch extraction mode may be a type of indicator extraction. In a switch extraction mode, processing the image data associated with the device interface source into an extracted interface representation can include parameterizing the mechanical state of a switch. The switch will preferably have at least two mechanically stable states, but could alternatively have multiple stable states. Multiple switches could additionally be monitored to determine some logical state. With switches and indicators, the change in state at times may be of interest in addition to or as an alternative to just the stable state.

In a dial extraction mode, processing the image data associated with the device interface source into an extracted interface representation can include parameterizing the settings of a one-dimensional dial. This functions to measure the position of a dial indicator relative to the operating range along one dimension. In a rotary dial this may include measuring the angle of the dial indicator. In a linear dial this may include measuring the position of the dial indicator along a linear path. Accordingly, this may include detecting the dial indicator and measuring position and/or orientation relative to the dial range. Dials can include dials that are set into a position by user input as well as dials that are set in response to some signal. In a rotating dial example, a

rotating dial can be translated to a base parameterized metric and then to a classified interface representation. A recognition stage of processing can include calculating an angle of the dial indicator, and a classification and data formatting stage can include translating the angle of the dial indicator to a signal value as shown in FIG. 36. A multi-dimensional dial or user input like a joystick with two-dimensions of positioning or another element could additionally be extracted.

The dial extraction mode can additionally be applied to many situations that don't involve a traditional dial such as treating the level of contents in a container or bottle as a dial indicator that moves linear along the path defined along the container.

In a character extraction mode, processing the image data associated with the device interface source into an extracted interface representation can include performing optical character recognition in a designated region.

Related to character extraction, the method could additionally include other forms of graphical interpretation that more generally includes parameterizing graphical state of image data of the device interface source, which functions to interpret graphical presentation of information in an interface output. In a basic variation, this may be through identification and classification of static or animated graphics. Such graphic classification can enable graphical interpretation of non-alphanumeric symbols such as logos, icons, infographics, symbols, and/or other visual objects or patterns. Parameterizing graphical state, in one variation, may be used in parameterizing a chart (e.g., a line graph, pie chart, bar graph, or other form of infographic) into a dataset. In the variation of interpreting a chart, interpreting the graphical representation may include identifying a graphic type, detecting dimensions, identifying at least one data-set indicator, and generating at least one data-set value through comparison of a data-set indicator to the dimensions according to the graphic type. Graphic type can specify the rules for how data-sets and dimensions are represented and how a data-set indicator should be evaluated according to the dimensions. A pie chart will have rules on angular interpretation of wedges, and a line chart will have rules on interpretation of a line plotted along a two-dimensional axis. In the case of a medical monitoring device, a biological signal reflected as a line graph can be translated into a time series data-set.

In a presence extraction mode, processing the image data associated with the device interface source into an extracted interface representation can include detecting at least one type of object in a region specified in the extraction configuration. Detection of an object can include various forms of CV-based object detection. When dealing with people or more uniquely identifiable objects, presence extraction can be used for unique identification of objects. In the case of people, facial recognition, and/or other forms of biometric identification may be used. Presence detection mode may additionally include counting objects, measuring the quantity, timing the duration of presence, or generating other metrics related to object presence. Multiple types of objects may be detected. In some cases, presence extraction is used in detecting presence of some combination of objects in a region. Presence detection may additionally be accompanied by physical state detection.

In a device model extraction mode, processing the image data associated with the device interface source into an extracted interface representation can include detecting device state of the device interface source, and processing the image data according to a detected device state. The



device model extraction mode functions to interpret device state of a device interface source. The device state is preferably detected according to a selected device model extraction mode set in the extraction configuration, wherein the selected device model extraction mode is on specifically for the particular device model of the device interface source. For example, each brand and model of heart monitoring device may have a different customized device model extraction mode to accommodate the different forms of data, device states, and device state transitions. Processing the image data according to a detected device state can include selecting a type of informational data for characterization based on the detected device state. Depending on the device state, different data (or lack of data) may be exposed through the interface. Processing the image data according to the detected device state can additionally include extracting a parameterized representation of the device interface source from the image data and characterizing the parameterized representation into a labeled, informational data.

Device state is a data characterization of the operating mode/machine state of a device and how that may alter the information exposed through a user interface. The device state may be reflected by what is being presented in the user interface but could additionally be reflected through other indicators such as input elements. The image data could be processed to detect the state of device inputs and other mode indicators such as indicator lights, physical device inputs, operating mode graphical labels. For example, the state of a set of switches, knobs, indicator lights may indicate what information is being displayed.

Device state is preferably interpreted according to a detected or identified device model as shown in FIG. 37. Complex devices such as medical devices, industrial control panels, and/or other device interface sources may be capable of displaying multiple pieces of information. Additionally, the displayed information may be modified based on the state of the device. For example, in one mode, a medical device may represent the heart rate as a large character representation of the current heart rate, but in a second mode, the medical device may represent the heart rate as a graph showing the heart activity signal used to calculate the heart rate. The logic of device inputs and other mode indicators can be characterized in a device model.

A device model may characterize the types of information displayed, when the information is presented, and other aspects of how the device outputs information. The device model could be identified automatically or through user identification. Automatic identification can use image classification. Additionally or alternatively, the brand, model number, a machine-readable code, or other suitable identifiers may be used to narrow the possible device models or to uniquely identify the device type. In another variation, the visuals of the device interface source may be used in identifying the device model.

The device state will generally change the type and format of information presented by the device interface source. The device state may be detectable based on a single snapshot image of the device but may alternatively depend on a monitoring historical transitions of device state and possibly interpreting a state machine governing state transitions. Detecting the device state preferably includes detecting a displayed mode indicator on the device interface source, tracking user interactions and device state, and/or other approaches. Mode indicators may include detecting state of physical elements of the device interface source such as extracting state of physical indicators like buttons, switches, LEDs, dials, and the like. Device indicators may addition-

ally include graphically displayed indicators such as GUI menu labels. In one variation, the method can include generating a device state model based on display images labeled as different device states. In this way small graphical styles of different device states (e.g., different applications, GUI widgets, and the like) can be efficiently learned and represented as a model for classification. For device interface sources with multiple applications, detecting device state may additionally include detecting an active application of the device interface source. Different applications, widgets or sub-programs may additionally have their own models for extraction, which can be activated when the application, widget, or sub-program is detected. In a particular device state information is preferably presented in different ways. Each device state may have different information that can be extracted. Maintaining a state machine of device state may additionally be used to associate information across different device states. For example, the method can interpret the navigation of a hierarchical menu in a modal operating system to understand classification of data extracted from a particular view.

With respect to device state detection, the method may additionally include detecting a user-object in the image data and tracking device state change in association with the user-object. The user-object is preferably a CV-detected person. In some cases, the user-object may be limited to specific human identities or classifications of humans (e.g., an worker, patient, customer, doctor, nurse, child, adult, etc.). People may often be the reason for a change in the operating state of a device interface source. Changes of a device interface source may be set to be conditional on the presence and/or interactions with a device interface source, which may simplify analysis and monitoring of device interface source. In other cases, detecting a user-object may be used in augmenting, supplementing or otherwise providing contextual data to data signals from a device interface source.

Block S134, which includes analyzing the extracted interface output functions to perform post processing on the extracted interface representation. Analysis may be performed on any suitable form of representation, but is preferably applied on parameterized, classified device interface representations (i.e., informational data). One or more data signals and/or device interface sources may be collectively analyzed. Depending on the application, different forms of analysis may be applied. More generally, basic analysis post-processing processes can be configurable such as averaging over a window, converting a rate of change or acceleration of a metric, or other basic processes. High-level logic and understanding of the information are preferably applied to interpret the extracted interface representations. For example, a heart rate signal extracted from a heart rate monitoring device may be analyzed to detect when the heart rate is normal or irregular. The analysis stage can be historical and/or real-time analysis. The analysis could be performed for a single source of information. An analysis stage may be used in generating alerts, triggering events, triggering interactions or commands with devices, systems, or services, and/or performing any suitable task.

In some variations, the analysis may additionally include integrating with one or more other data sources as a form of integration stage, which may enable more in-depth analysis. In an integration stage, multiple sources of information can be combined. The integration stage could be optional if only one device interface source is monitored. The integration stage could additionally combine with other information sources such as other sensing systems of similar or different



types, or information systems. The integration stage may organize or restructure the extracted interface representations for easier consumption and/or monitoring.

The various types of extraction can additionally include other processing features to account for different operating attributes of a device interface source. Indicator interfaces may also communicate information through a sequence of blinks. These sequences could be configured for detection, and block S130 could automatically detect these blinking conditions. Dial devices may offer the ability to move the dial continuously or more than a full rotation. Processing of a dial that is configured indicating such rotation range may be specifically tracked to account for being rotated beyond. Continuous tracking of the indicator may be used. Character based devices may operate with varying mechanics. A character-based display may scroll horizontally, scroll vertically, or cycle through messages. Such display changes could be automatically tracked and accounted for in the processing stage to create better-structured data. Other device mechanics could additionally be accounted for.

Additionally, during execution of the method, there may be gaps in the monitoring a device interface source. The device interface source may be undetectable because the view is blocked, the lighting conditions change, the device interface source is temporarily moved, or other complications. The method can preferably account for inconsistent data collection. The reason for a capture status change could be classified. Collected extracted interface representation data could be annotated indicating the capture status. Additionally, configuration for extraction of device interface source could be set to send a notification or issue an alert during particular capture status conditions.

As discussed, the method may additionally be usable in the detection and interpretation of audio interface elements. In a variation applying audio interface extraction, the method can include collecting audio data and detecting audio output of a device interface source and characterizing an extracted interface representation of the audio output.

The audio data can preferably be processed into an informational data that represents the meaning or informational content of the audio. Audio signals are commonly used to indicate state. For example, different beeps or alarms can signal different pieces of information. In a preferred implementation, the collected image data is audio-visual data where audio is recorded in synchronization with visual images. Audio may alternatively be collected independently and in some cases independent of image data. In one example, a device interface source may have one or more alarms that are sounded under different conditions. The device interface source could be configured to detect that alarm. As a data signal this, in one example, could be represented an event record when the alarm is detected.

Block S140, which includes exposing at least one access interface to the extracted interface representation, functions to utilize the information obtained from the devices. The method could be applied in a variety of use cases, and the type and manner of exposing an interface could be different depending on the use case. The exposed interface could be a user interface or a programmatic interface. The user interface could be accessed via a native application, a web application, or any suitable type of graphical, visual, or auditory user interface. Additionally, the user interface could be facilitated through a communication channel. Notifications and alerts could be triggered based on different conditions of the extracted interface representations. For example, an alert could be sent to one or more recipients in response to the classified interface representation satisfying

some condition (e.g., heart rate dropping below a set value). A programmatic interface could be an application programming interface, a data communication protocol, a data storage solution, application processing routine, and/or any suitable mechanism enabling programmatic interactions. The programmatic interface can enable other applications and services to integrate with the interface extraction system.

The extracted interface representation data is preferably organized or structured in a consistent manner within the interface extraction system. In one variation, the extracted interface representation data could be presented with different elements of information prioritized, summarized, or presented so that a user is initially presented with the most relevant or critical information. The extracted interface representation of one or more device interface sources can be queryable and sortable.

A programmatic interface to the extracted interface representation(s) can be used to build various sets of tools and/or services. A programmatic interface may include storing the extracted interface representation and providing API access. Exposing at least one access interface to the extracted interface representation could include exposing an API to a historical data record of the information data. In another variation, the extracted interface representation could be directly communicated to another system or application. For example, a digital assistant, electronic product, or other form of device or product could be directly sent the extracted interface representation. In some variations, the extracted interface representation can be converted to a control line so that the state of another connected system can be controlled in response to the extracted interface representation.

In another variation, exposing an interface can include streaming a rendered interface representation. In one implementation, the rendered interface representation is formatted image data prior to data extraction. Some embodiments may do this independent of data extraction, where basic identification, cropping, and image formatting can be used to provide video surveillance of device interfaces. As shown in FIG. 35, video and/or images of various device interface sources may be formatted into a rendered interface representation and streamed for easier browsing and remote viewing.

More generally any suitable form of extracted interface representation may be streamed. The data could be in a parameterized interface representation, a classified interface representation, or data from higher-level analysis. For example, information data could be streamed as a real-time data stream. In another variation, the rendered interface representation can be a virtual rendering of the interface that is driven by data extracted from the device interface source. In one implementation, the rendered interface representation can be a digitally assembled representation of the device interface source using informational data. This may provide a similar user experience to video surveillance of a device interface source, but presenting formatted extracted data instead of image data. This may be done so that a common graphical presentation of information can be presented to a user despite possible differences in the make and model of the underlying device interface source of the data.

The streamed rendered interface representation can be used in remote surveillance and, in some implementations, may be streamed to a remote monitoring system. The remote monitoring system could be accessible as a user interface. A user interface of the interface extraction system can be provided to organize one or more device interface sources



for remote monitoring. The rendered interface representation can additionally be streamed and/or accessed by multiple distinct parties. In one variation, the rendered interface representation includes at least part of the image data. The image data is preferably processed by at least a formatting stage. Intelligent streaming could include only transmitting changes of the device interface source.

The remote monitoring system is preferably configured to manage a set of media streams of formatted image data or other forms of extracted interface representations from device interface sources.

When applied in an industrial application where multiple device interface sources may benefit from grouped human monitoring, a rendered interface representation preferably enables the monitoring of a plurality of device interface sources. The various device interface sources may be organized or categorized for group monitoring and/or analysis. The device interface sources could be grouped within the interface using some logical grouping such as environment location, patient/worker association, or any suitable grouping. Additionally, the visual or data representation of one or more device interface sources can be presented in a structured format that organizes and prioritizes appropriate information. In one exemplary use case, rendered interface representations grouped by room and/or patient can be streamed to a head nurse overseeing a floor of patients in a hospital. The various device interface sources for each patient are grouped together. The head nurse can have an overview view where all patients can be monitored within one view. A patient view can enable the nurse to select a particular patient for detailed monitoring—the history of the device interface sources and/or lower priority information may be presented. The data may additionally be grouped by other assigned hospital staff members so that doctors, nurses, floor managers and the like could be provided with an interface for quickly reviewing relevant groups of data. This could be similarly applied in other industries and applications.

The method may additionally include selectively elevating priority of a device interface representation. This is preferably used when monitoring multiple device interface sources and the priority is set according to an analysis result involving the device interface representation. In the case of a remote monitoring system, this may be applied to changing the priority of a rendered interface representation according to the informational data. This could be particularly useful in exposing a rendered interface representation for surveillance and monitoring purposes. Analysis could be applied on the extracted interface representation data and determine a priority. In a hospital this may include prioritizing data relating to patients that have signals outside of normal conditions. This could be used to self-organize the presented device interface source information. In the example of the head nurse above, the remote monitoring system may highlight the display of nine active rooms at one time, but those active rooms may be dynamically changed according to priority levels. This could additionally be combined with normal video surveillance. A video stream of some location may have elevated priority based on extracted data from device interface sources. In a manufacturing plant, a remote monitoring system used for video surveillance monitoring may automatically select displaying a video stream monitoring a first device when the extracted interface representation of the first device moves outside of normal operating conditions.

In one implementation, the method may be used in combination with a programmatic event-handling system.

The event-handling system may function to process the extracted interface representation and trigger a programmatic event upon detecting an event. Various event conditions could be monitored. The event conditions may be based in part on some aspect of the extracted interface representations. Then upon satisfying the condition a programmatic event could be initiated. In another implementation, the interface to the extracted interface representation(s) could be integrated with a home automation hub wherein various IoT devices could be collectively managed. In another implementation, a user interface for remotely monitoring the device interface source could be created. A wide variety of applications and tools could be constructed using the method.

The method may additionally include monitoring the capture status of a device interface source and generating feedback in response to a change of capture status of the device interface source **S150**, which functions to detect and respond to problems of a device interface source being lost from view, blocked, angled to prevent accurate extraction, or otherwise obscured in the field of view of the imaging device. The feedback is preferably applied to address the changes so that actions can be taken to resolve issues.

The capture status relates to the state of image data associated with device interface source that is used in processing. The capture status could be a binary status indicating if an extracted format can be generated. In another variation, the capture status could be a reliability metric. Under ideal lighting conditions and a clear view of the device interface source the reliability metric can have a high rating. As the lighting conditions become less ideal and/or the angle of the device interface source moves off center, the reliability metric may decline. At some level the capture status can indicate that a reliable extracted interface representation is not obtainable.

Audio interface elements could additionally have a capture status. Capture status may be impacted by background noise, the volume settings of the device, the orientation of a microphone of the interface extraction system, and/or other properties.

Generating feedback in block **S150** can be implemented in a variety of approaches. In one approach, the imaging device or another suitable device could generate user feedback. The user feedback could be an audio alert such as a sound or announcement. The user feedback could additionally or alternatively be a visual alert such as activating a status light. The user feedback can communicate the issues with the capture status. For example, an audio announcement may announce that the brightness settings of the display on a device need to be changed. In another example, a graphical display on an imaging device may indicate the “signal strength” based on the image data conditions.

In one preferred implementation, generating feedback can include announcing a change in capture status during user exit of the field of view. In this implementation, the method may additionally include tracking user presence in the field of view of the image data, which can enable detecting the user exit. In some cases, humans may commonly enter the field of view obscuring or altering the capture status of a device interface. While a human that causes the change is present in the field of view, the method can assume that the human is interacting with the device interface sources and performing the task of monitoring information from the device interface source. In other words, the method can temporarily delegate device interface source monitoring to a user when the user is in the vicinity of the device interface source. This can be particularly useful in a hospital use case



where nurses and doctors will enter a patient's room, view the device interface sources, and interact with the device interface sources. The method may additionally be able to distinguish between different users. For example, the method can be configured to distinguish between hospital workers and other people (e.g., patients, patient family members, etc.). The appropriate users can be tracked through the image data (or other sources of location information like GPS or RF-based triangulation), and when the user leaves the proximity of the device interface source, the user can be alerted to any issues with the monitoring conditions.

As an exemplary scenario shown in FIG. 38, a nurse walks into a patient's room. While the nurse is checking on the patient, they may move a monitor device and/or block the view of the monitor device. During this time the capture status will indicate that data cannot be collected from the device interface source, but user feedback is not generated while the nurse is still helping the patient. When the nurse walks out of the room and is detected to leave the field of view of the image data, an audio announcement could play if the capture state of one or more device interface sources needs adjustments. The nurse will ideally hear and respond to the audio announcement and adjust the various device interface sources or make other adjustments to address the capture state issues.

Feedback may alternatively be generated and associated with data of the extracted interface representations. In cases where a device interface source is lost, the system can accommodate such gaps in information by flagging or annotating data. In some cases, the capture status could be a parameter associated with all extracted data formats. Similarly, the manner in which a device interface source was lost could be classified and used to flag or annotate the data. Classifications for lost device interface sources can include obscured device, lighting conditions, off-angle view, and device interface source disappearance.

To address possible changes in the capture status, the imaging device could be in close proximity to a device interface source to mitigate the chances of an object blocking view of the device interface source. In one variation, the imaging device can be mounted directly to the device interface source such that the relative position and orientation of the imaging device and the device interface source is fixed.

Within a system integration, the method is preferably configured to provide detection of two or more device interface sources, processes the image data into extracted interface representations for each of the two or more device interface sources, and exposing an access interface to the extracted interface representations. The two or more device interface sources could be a combination of the same or different types of device interface sources. They may also be observed and detected in the image data from the same or a combination of imaging devices. When used for different device interface sources, detecting a set of device interface sources observed in the image data preferably includes detecting a type of each of the set of device interface sources and then selectively processing the image data of each device interface source based on the type of device interface source. For example, for one instance of a device interface source processing the device interface source may include performing character recognition, and for a second instance of a device interface source, processing the device interface source may include parameterizing graphical state of a

visual output of the device interface source. As discussed, the method could additionally be used for a single device interface source.

In one implementation, multiple imaging devices may be used to collect extracted interface representations of a device interface source. These different imaging devices may be used at distinct times or overlapping times. In one exemplary scenario, image data is captured by at least a first imaging device and a second imaging device; and processing the image data associated with the device interface source into an extracted interface representation can include processing image data collected from the first imaging device at a first instance into an extracted interface representation of the device interface source during the first instance and processing image data collected from the second imaging device at a second instance into an extracted interface representation of the device interface source during the second instance.

The method is preferably operable for use with multiple device interface sources, wherein the method may include setting two or more device interface sources, and for each instance of extraction configuration of a device interface source, processing the image data associated with the respective device interface sources into extracted interface representations according to the respective extraction configuration. The extracted interface representations of the different device interface sources are preferably managed as separate and distinct data feeds (unless configured for merging). The exposed access interface may be used to access the extracted interface representations of select device interface sources or multiple device interface sources.

In one variation of the method, the image data is collected by a movable imaging device. A movable imaging device may be a camera on a smart phone, a tablet computer, a wearable computer such as smart glasses or a body camera, an imaging system of a robot, automobile, or movable machine, or any suitable device that is generally not statically positioned.

The movable cameras will preferably be exposed to many more potential device interface sources. And the method may employ automatic enrollment and configuration of device interface sources to facilitate accommodating the volume of device interface sources to be extracted. In one variation, the device interface sources may be automatically configured on demand in response to some condition. In another variation, the method can include enabling proactive processing of detected device interface sources that are pre-configured with extraction configuration.

In a movable imaging device variation, the method may include collecting image data S110; setting extraction configuration of a device interface source identified in the image data, where the extraction configuration includes setting a device position S120; tracking an image collection position S160; processing the image data associated with the device interface source into an extracted interface representation according to the extraction configuration and responsive to when the image collection position corresponds to a configured device position of the device interface source S135; and exposing at least one access interface to the extracted interface representation S140 as shown in FIG. 39.

Block S160, which includes tracking image collection position, functions to detect position and/or orientation within an environment or global setting. In one variation, tracking image collection position includes collecting global positioning coordinates and/or device orientation measurements of the movable imaging device as shown in FIG. 40. In one implementation, the imaging device is integrated into a computing device with a location service that can leverage



GPS and/or cellular/Wi-Fi location detection. The computing device may additionally include an inertial measurement unit (IMU), which may include an accelerometer, a gyroscope, and a magnetometer. The IMU may be used to detect the orientation of the imaging device and direction of the imaging device. The image collection position may additionally include determining a spatial estimation of an environment from collected image data. A spatial estimation can use various spatial sensing technology integrated into the computing device, processing multiple images, and/or using other techniques used in generating a spatial map. The spatial map may additionally be used in understanding what is viewable by the movable device at different moments.

Additionally or alternatively, tracking image collection position can include tracking visual mapping of a scene, which functions to use image mapping within a localized area. The image data can be compared to past image data of a device interface source. The visual mapping will signal when similar scenes are detected which can increase the likelihood that the device interface source can be detected. Visual mapping may be activated when GPS indicates approximate proximity to a device interface source.

In one variation, tracking of image collection position can be used in setting the device position during setting of the extraction configuration. During setting of the extraction configuration, positioning of the movable imaging device is preferably used to establish an estimation of the location of a device interface source. GPS and location information can provide rough location. Orientation may provide information such as if the device interface source is located high or low. Spatial estimation can be used to understand from where the device interface source may be viewable.

Tracking of image collection position is additionally used in determining when image data should be processed. The method can search for device interface sources that may be viewable based on the tracked image collection position. It can similarly be applied to a selectively activate collection of image data. For example, a device may initiate collection of image data in response to the device position corresponding to a possible device interface source.

In a glasses variation, eye tracking may additionally be used in identifying the object or region of image data viewed by a user. This may be used in directing the processing of image data to generate an extracted interface representation. This may additionally be used in selecting an object for extraction configuration.

The device interface sources may be associated with the device and user, but a shared platform may additionally collect device interface sources so that the network of imaging devices can cooperate in sharing extraction configuration and collection image data and processing image data.

The method may additionally include detecting a candidate device interface source. In one variation, such automatic enrollment and configuration can be performed independent of any configured device position to enable on-demand extraction of data from a device interface source. The method actively captures image data and processes the image data for device interface source detection upon detection. Interface source detection can use CV-based object classification and detection techniques. When a device interface source is detected, it can be processed according to an extraction configuration based the type of device interface source. Without configuring a device position the extracted interface representation may not be part of a larger historical record of the device interface source and instead provide a snapshot for the duration that device interface source was

viewable. Alternatively, detection of a candidate device interface source can be used along with setting device position such that a historical record of that device interface source can be generated. Accordingly, the method shown in FIG. 39 can include detecting at least one type of device interface and automatically setting extraction configuration of a device interface upon detection of a type of device interface, wherein the extracting configuration sets device position to map to the image collection position.

The movable camera variation of the method may include the scenario where multiple imaging devices that are operated by different users can collect extracted interface representations of a device interface source at different times. The method may enable sharing of extracted interface representations so that the collaborative collection of imaging data and processing can provide more data points of a device interface source. Sharing can function to share data across accounts and devices. Sharing can be particularly useful for a movable camera, as image data of a device interface source will likely only be periodically processed. Sharing of data and processing may expand the number of opportunities to observe and process a device interface source. For example, a family with sharing of extraction configuration for device interfaces at their home may be able to provide more data on monitored devices. An interface extraction platform in communication with the different imaging devices can preferably coordinate the sharing of image data and/or extracted interface representations.

There may be cases though where one would desire to restrict capabilities to perform interface extraction. The movable camera variation (along with other variations) may enable permissions to augment the sharing of image data and/or extracted interface representations. A policy engine can preferably enable, disable, or regulate processing of image data into extracted interface representations and the accessing of generated extracted interface representations of different devices based on the privileges associated with that account. The interface extraction platform in communication with the devices may be able to remotely enforce the policy engine. Alternatively, the policy engine may have access to a local copy of permissions that can be enforced on-device.

In one variation, imaging devices can be associated with user accounts such that registering of an imaging device can set how permissions are enforced for that imaging device. The user accounts may additionally be used in accessing data or services using the extracted interface representations. Any suitable identity mechanism may be used in place of user accounts. In one instance, a movable camera may be associated with a first account and used in collecting image data that is processed into an extracted interface representation. However, exposing an access interface to the data may permit a second account to access the extracted interface representation and prohibit and/or restrict the first account from reading, accessing, or otherwise using the extracted interface representation.

The permissions can permit, restrict, or otherwise limit the processing of image data into extracted interface representations and/or accessing collected extracted interface representations. Permissions for processing can be used to prevent conversion of image data into an extracted interface representation. Permissions for accessing may be used to allow the collection of image data for processing but then prevent access from particular entities possibly including the user that assisted in collecting and/or processing the image data. In one instance, the permissions can be used to grant one particular type of account permission to facilitate the



generation of extracted interface representation data but to prevent access to the data the account generated. In an example of this instance, a first user account is associated with an imaging device and is permitted to access extracted interface representation data for a particular device interface source. In this example, a second user account is associated with a second imaging device may also assist in the generation of extracted interface representation data, but the second user account may be prohibited from accessing the extracted interface representation of the device interface source through an exposed interface.

Permissions can be conditional on associated account, device position, device interface source identity (e.g., a particular device interface source), device interface source type (e.g., a particular type of device source such as a particular type of computing device), time of day, number of data points, and/or other factors. Permissions can be set for specific user accounts or based on properties of user accounts such as group association (e.g., part of a family, friend, or work group), residency, and the like. Position and geographic region may be used in many cases for conditional permissions.

Permissions may be set by administrators of the system. Permissions may alternatively be set by individual users. In one implementation, the method can include granting permission administrator rights to a restricted geographic area to a user account. Accounts could alternatively be granted permission administrator rights over particular device interface source types (e.g., manufacturers of a particular device). Claiming of permission administrator rights preferably includes validating the request to ensure the account should be granted those rights. For example, granting permission administrator rights to a restricted geographic area may depend on proof of residency and/or ownership of that geographic area. In this way, homeowners and/or business owners can restrict and control interface extraction with and/or at their property. In one instance of granting permission administrator rights, the method may include setting regional permissions of extraction configuration to a first account for a first region; wherein setting of the extraction configuration of a device interface source is restricted to configuration by the first account. The first account or set of accounts in this example, is the only one enabled to set extraction configuration for a device interface source in that first region. Alternative instances may allow some limited options of setting extraction configuration. For example, a second user that brings an object already associated with extraction configuration into the first region may still be permitted to manage permissions for that object.

#### 4. System

As shown in FIG. 41, a system for establishing remote integration with a device through a user interface output of a preferred embodiment can include an imaging system 110, a device interface processing engine 120, and an access interface 130. The interface extraction system is preferably used for extracting user interface outputs of a device that would be available to a user for interpretation and converting the related information to an extracted interface representation for remote access and integration. The device interface source preferably includes a visual interface but may additionally or alternatively have audio-based interface elements. The user interface of a device can be an explicit user interface like a display but may alternatively be an analog user interface like the fluid level in a container (e.g., an I.V. bag level), a knob position, or dial position. The system preferably implements the method described above but may alternatively be used in any suitable manner.

The system can be implemented to monitor a single region. The system may alternatively be implemented to monitor a plurality of distinct regions. For example, a hospital may have multiple imaging systems 110 installed in multiple patient rooms, and the system can facilitate individually monitoring the various device interface sources in the multiple patient rooms.

In one embodiment, the system is implemented within a remote interface extraction platform. The interface extraction platform can integrate with the imaging system no, include at least a portion of the device interface processing engine 120, and provide one or more access interfaces 130. In one variation, the remote interface extraction platform is a multi-tenant platform wherein multiple different accounts can run distinct instances of the system through the platform. Alternatively, a single-tenant solution could be used in place of a remote interface extraction platform. For example, a server application could act as an on-premise platform. In yet other implementations, a stand-alone imaging device may be designed to include the imaging system 110, device interface processing engine 120, and the capture device state interface.

The imaging system 110 functions to collect media and more preferably the image data that may contain a device interface source. The imaging system no can include one or more imaging devices. The image data collected by the imaging system is preferably video but can alternatively be a set of periodic static images. The imaging devices are preferably visual video or still cameras, but may additionally or alternatively collect infrared, depth-based, lidar, radar, sonar, and/or other types of imagery. The imaging system no can additionally include a low-light/night mode. For example, an IR illumination system could emit IR light and collect image data from the IR spectrum during low light or nighttime mode. Depth imaging devices, and other suitable forms of imaging devices may additionally or alternatively be used. An imaging device is preferably mounted such that a region of interest is in the field of view of the imaging device. The imaging device can be static such that the field of view is substantially not dynamic. Alternatively, the imaging device could be actuated where the field of view can be changed by rotating, elevating, panning, zooming, moving and/or otherwise altering the field of view. In one variation, the imaging system 110 could include a free moving imaging device (i.e., a movable imaging device) such as one worn by a user. An imaging device is preferably a standalone imaging device with a power source and a network interface (e.g., wired or wireless). Processing may be performed on the imaging device but may alternatively be performed entirely or in part on a remote device. In one variation, the imaging system 110 could be a vision system interface that enables integration with an existing imaging system no (e.g., surveillance cameras).

In a variation with a movable imaging device, the imaging device can additionally include or have access to positioning and/or orientation sensing devices such as a location service of a computable device (e.g., using GPS and/or cellular/Wi-Fi location) and/or an IMU (e.g., including a three-axis accelerometer, a three-axis gyroscope, and a magnetometer).

In a variation with an imaging system network, the imaging system 110 can include a multitude of imaging devices distributed in the environment with the imaging devices positioned at a range of distinct vantage points. When installed for enabling interface extraction across an environment, the imaging system 110 may be a high-density imaging system. A high-density imaging system is preferably characterized by a large portion of the relevant portions



of environment normally observed by an image capture device. A large portion, in one example, can be characterized as greater than 95% of surface area of interest. High density may additionally be characterized as having redundant coverage. In one example, high density imaging system may be characterized by one camera for every one hundred square feet of surface area of interest (e.g., the ground, product storage faces, etc.). In an environment like a small grocery store this can may be twenty or more cameras distributed for coverage of two thousand square feet of surface area of interest. The imaging device to environment space ratio could be any suitable ratio. The system could alternatively use a sparse network of imaging devices with little or no overlap of fields of view.

The imaging system **110** preferably continuously or periodically collects image data that is processed by the device interface processing engine **120**. The imaging system no may alternatively be dynamically controlled to collect image data on demand to collect image data for the device interface processing engine. The capture configuration of an imaging device could be statically set but may alternatively be dynamic. Capture configuration can include any suitable imaging setting such as ISO, aperture, shutter speed, zoom, or other settings. The capture configuration could be dynamically altered based on one or more results of the device interface processing engine **120**. The capture configuration could additionally operate in alternating capture configurations so as to cycle through camera settings customized for different device interface sources.

Herein, the system is primarily described as it would be used for visual interface extraction. The system could additionally or alternatively utilize one or more microphones to collect audio, vibration, and ultrasonic signals and convert the collected data into an extracted sound-pressure interface representation of the device. The system could additionally or alternatively utilize one or more chemical sensors to collect signals and convert the collected data into an extracted chemical interface representation of the device. In an alternative embodiment the system can be implemented with only visual, only microphones, or only chemical sensors, or any combination where interface cues can be collected and processed without other forms data collection.

The device interface processing engine **120** functions to process the image data from the imaging system to extract the state of a device interface source. The device interface processing engine **120** is preferably configured in cooperation with the imaging system no to execute the processing of image data into an extracted interface representation described above. In one implementation, the device interface processing engine **120** can include a pre-processing model and/or an analysis engine. The device interface processing engine **120** preferably operates on image data associated with a device interface source. The region of image data associated with a device interface source could be pre-configured but may alternatively be automatically detected.

The device interface source processing engine can include a device detection module, a pre-processing model, and a device state interpreter. One or more analysis engines could additionally be integrated with the processing engine.

The device detection module functions to identify, track, and extract image data for a device interface source. The device detection module preferably dynamically detects the relevant visual interface components of a device. The device detection module can be pre-configured with a number of different device models that can be automatically recognized and set to an associated extraction configuration. For example, the system can be configured to automatically

recognize and initiate interface extraction for over one hundred different medical devices. The device detection module can handle automatic enrollment (and disenrollment) of a device as it enters (or leaves) a field of view. Automatic enrollment functions to accommodate dynamic environments like a patient room where the device interface sources can frequently change. In some variations, the interface of a device interface source may be extracted without specifically identifying the type of device. For example, a graphical display may be detected and automatically extracted and processed. The device detection module may additionally or alternatively have a manual selection mechanism where a user can setup the extraction of a device interface source.

The device detection module can additionally extract multiple device interface sources from one field of view. For example, one imaging device may be able to view multiple devices and information can be extracted from each device.

The pre-processing module functions to transform image data that relates to a device interface source. A cropped copy of image data can be generated for each device interface source. Various visual transformations including geometric transformations and imaging transformations can be applied to the image data within the pre-processing module.

The analysis engine functions to synthesize image data associated with a device interface source into an interpreted representation. There could be multiple stages and processes executed by the analysis engine. The analysis engine preferably operates on the output of the pre-processing module.

The analysis engine could facilitate processing a classification stage, an integration stage, and/or one or more higher-level analysis stages. In a classification stage, optical character recognition can be applied on the image data. Image/symbol classification, color detection, audio pattern matching, and/or other forms of classification can be used to generate a parameterized interface representation. A secondary classification stage can operate directly on the parameterized interface representation to interpret the presented information. Additionally, the device state interpreter can be used to interpret the presented information. An integration stage functions to combine multiple pieces of information. The different pieces of information can be retrieved from one or more device interface sources. Various higher-level analysis systems can perform real-time or historical analysis of the generated information. For example, higher-level analysis could analyze all the different meters and initiate an alert.

The device state interpreter functions to apply a device model to determine the symbolic meaning of the interface output of a device interface source. The device detection module can generate a mapping between a device interface source and a device model. The device models could characterize, classify or provide guidance on how to process a particular device interface source. The device model may characterize the layout of information. A device may display multiple pieces of information simultaneously, and the device model could define what regions of the display are associated with what type of information. A device model may additionally include operating state characterizations of the device. The device state interpreter can utilize the operating state characterizations to interpret displayed information when the device may display a variety of information depending on how it is configured.

The access interface **130** functions to offer some form of access to the collected extracted interface representations. The access interface **130** could be a user interface such as a dashboard for accessing and exploring collected data. The



access interface **130** could alternatively be a media communication interface such as a formatted video stream or notification/alert system. The access interface **130** could alternatively be a programmatic interface such as an application programming interface (API) or an application processing engine. An API can enable outside applications or services to remotely interact with the collected data. An application processing engine could enable scripts or applications to be processed and executed at least in part by the system. For example, higher-level analysis processes may be designed and configured by a third party.

The system could additionally include a control interface, which functions to enable configuration of the imaging system and/or the device interface processing engine **120**. The control interface is preferably a user application operable on a computing device such as a smart phone, a personal computer, a wearable computer, and/or any suitable computing device. When setting up the system for a device interface source, the user can use the control interface to set various configuration options. The user could specify the position of the device interface source within the image data, the type of extraction, set some of the operating properties. In one variation, live image data can be streamed from the imaging system no to the control interface so that configuration can be performed and previewed on the current state of the device interface source. Additionally, the control interface may be one user interface for viewing and monitoring extracted interface representations of the device interface source.

The system may additionally include an interface extraction platform that is a network accessible platform or service that can be used to facilitate in the processing and/or management of data.

The interface extraction platform in one implementation can host the device interface processing engine **120** in part or full. A control interface like a user application may have communication access to the interface extraction platform so that processing tasks and/or data can be synchronized between different device instances. APIs, web dashboards, administrator control panels, and/or other forms of access interfaces may be provided through the interface extraction platform. In one variation, extraction configuration and data on the various device interfaces can be stored and managed by the interface extraction platform.

The system may additionally include a policy engine that functions to manage sharing of data and interface extraction capabilities. The permissions that can be set by the policy engine are preferably substantially similar to the ones described above. The policy engine can be integrated into the interface extraction platform but may alternatively operate locally with the device interface processing engine **120**.

The systems and methods of the embodiments can be embodied and/or implemented at least in part as a machine configured to receive a computer-readable medium storing computer-readable instructions. The instructions can be executed by computer-executable components integrated with the application, applet, host, server, network, website, communication service, communication interface, hardware/firmware/software elements of a user computer or mobile device, wristband, smartphone, or any suitable combination thereof. Other systems and methods of the embodiment can be embodied and/or implemented at least in part as a machine configured to receive a computer-readable medium storing computer-readable instructions. The instructions can be executed by computer-executable components integrated with apparatuses and networks of the type

described above. The computer-readable medium can be stored on any suitable computer readable media such as RAMs, ROMs, flash memory, EEPROMs, optical devices (CD, DVD, etc.), hard drives, SSHDs, or any suitable device. The computer-executable component can be a processor, but any suitable hardware device can (alternatively or additionally) execute the instructions.

As a person skilled in the art will recognize from the previous detailed description and from the figures and claims, modifications and changes can be made to the embodiments of the invention without departing from the scope of this invention as defined in the following claims.

We claim:

**1.** A method for uncoupled remote device monitoring comprising:

collecting image data;

detecting a device interface source present in the field of view of the image data;

extracting a parameterized representation from image data of the device interface source;

detecting device state;

characterizing an extracted interface representation from the parameterized representation, the characterization being a process based on the device state; and

exposing at least one access interface to the extracted interface representation.

**2.** The method of claim **1**, wherein detecting device state comprises detecting a displayed mode indicator on the device interface source.

**3.** The method of claim **1**, wherein detecting device state further comprises detecting state of physical elements of the device interface source.

**4.** The method of claim **1**, wherein detecting device state comprises detecting an active application of the device interface source.

**5.** The method of claim **1**, wherein detecting device state comprises detecting audio indicating state of the device interface source.

**6.** The method of claim **1**, further comprising detecting a user-object in the image data and tracking machine state changes in association with the user-object.

**7.** The method of claim **1**, further comprising analyzing data patterns of the extracted interface representation.

**8.** The method of claim **1**, collecting image data comprises collecting audio-visual data and wherein characterizing the extracted interface representation from the parameterized representation further comprises detecting audio output of the device interface source and characterizing the extracted interface representation in part from the detected audio output.

**9.** The method of claim **1**, wherein characterizing the extracted interface representation from the parameterized representation comprises characterizing at least two sets of information data, used in the extracted interface representation, from the parameterized representation.

**10.** The method of claim **1**, wherein extracting the parameterized representation of the image data of the device interface source comprises performing character recognition in at least one region and parameterizing graphical state of a second region of the image data.

**11.** The method of claim **1**, wherein exposing at least one access interface to the extracted interface representation comprises streaming a rendered interface representation to a remote monitoring system, and further comprising selectively elevating priority of the rendered interface representation according to informational data in the rendered interface representation.



49

12. The method of claim 1, exposing at least one access interface to the extracted interface representation comprises exposing an application programming interface to a historical data record of information data from the extracted interface representation.

13. The method of claim 1, wherein the device interface source is a scale, and the extracted interface representation includes measured weight.

14. The method of claim 1, wherein device state is conditional on detected user interactions with the device interface source.

15. The method of claim 1, further comprising:

detecting a second device interface source present in the field of view of the image data;

processing the image data associated with the second device interface source into a second extracted interface representation; and

wherein the second extracted interface representation is exposed through the at least one access interface.

16. The method of claim 1, further comprising monitoring the capture status of a device interface source and generating feedback in response to a change of capture status of the device interface source.

17. The method of claim 16, further comprising tracking user presence in the field of view of the image data and wherein generating feedback in response to a change of the capture status of the device interface source comprises announcing a change in capture status during user exit of the field of view.

18. A non-transitory computer-readable medium storing instructions that, when executed by one or more computer processors of a computing platform, cause the computing platform to:

50

collecting image data;

detecting a device interface source present in the field of view of the image data;

extracting a parameterized representation from image data of the device interface source;

detecting device state;

characterizing an extracted interface representation from the parameterized representation, the characterization being a process based on the device state; and

exposing at least one access interface to the extracted interface representation.

19. A system comprising of:

one or more computer-readable mediums storing instructions that, when executed by the one or more computer processors, cause a computing platform to perform operations comprising:

collecting image data;

detecting a device interface source present in the field of view of the image data;

extracting a parameterized representation from image data of the device interface source;

detecting device state;

characterizing an extracted interface representation from the parameterized representation, the characterization being a process based on the device state; and

exposing at least one access interface to the extracted interface representation.

\* \* \* \* \*